Lecture 8

Topics:

I) The conditional distribution of T given the data d & C_ℓ^{TT}

II) Wiener filtering.

We have 2 equivalent ways of writing our idealized data model

Pixel domain) $d(\hat n) = \underbrace{\varphi * T(\hat n)} + \sigma \underbrace{w(\hat n)}_{\text{unit white noise}}$

$\varphi_{\ell m} = \frac{1}{2\pi}\sqrt{\frac{2\ell+1}{4\pi}} \exp\left(-\frac{b^2}{8\log 2}\ell(\ell+1)\right)$

observed on all $\hat n \in S^2 \subset \mathbb{R}^3$.

Harmonic domain) $d_{\ell m} = T_{\ell m} + \underbrace{\varepsilon_{\ell m}}$

has spectral density $\sigma^2 \exp\left(\frac{b^2}{8\log 2}\ell(\ell+1)\right)$

observed for all $\ell = 0, 1, \dots, -\ell \le m \le \ell$

The previous lectures derived the posteriors

$$P(C_\ell^{TT} \mid d) \quad \& \quad P(\theta \mid d)$$

where $\theta = (\Omega_c h^2, \Omega_b h^2, \tau, \theta_s, A_s, n_s)$.

However, the raw temperature map $T(\hat n)$ is also of scientific interest. In this lecture we will analyze the map posterior (assuming C_ℓ^{TT} is known)

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

← bandpowers are no longer sufficient statistics

After this we will combine the two to study the full posterior

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

Finite dimensional Analog

Suppose $d = s + n$ ← think of this as the analog to the pixel domain model.

data vector, signal vector, noise vector

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

$n \sim N(0, N)$

are independent where Σ & N are known.

Then the conditional density of s given d is

$$p(s \mid d) \sim N\Big(\underbrace{E(s \mid d)}_{\text{conditional expected value}}, \underbrace{\operatorname{var}(s \mid d)}_{\text{conditional var/cov matrix.}}\Big).$$

A fancy way to derive $E(s|d)$ & $\operatorname{var}(s|d)$ is to notice that

$$E(s|d) = \operatorname*{argmax}_s \log P(s|d).$$

$$= \operatorname*{argmax}_s \left[\log P(s,d) - \log P(d)\right]$$

$$= \operatorname*{argmax}_s \log P(s,d).$$

& similarly

$$\operatorname{var}(s|d)^{-1} = -\nabla_s^2 \log P(s|d)\Big|_{s=E(s|d)}.$$

$$= \underbrace{-\nabla_s^2 \log P(s,d)\Big|_{s=E(s|d)}}_{\text{Fisher information.}}$$

(Note: we don't need an expected value here since this hessian doesn't depend on s or d)

Now for $E(s|d)$ find the stationary points (i.e. $\nabla_s \log p(s,d) = 0$)

$$\nabla_s \log p(s,d) = \nabla_s \left[\log p(d|s) + \log p(s)\right]$$

$$= -\nabla_s \left[\frac{(d-s)^T N^{-1}(d-s)}{2} + \frac{s^T \Sigma^{-1} s}{2}\right]$$

$$= -\left[-N^{-1}(d-s) + \Sigma^{-1} s\right]$$

$$= N^{-1} d - (N^{-1} + \Sigma^{-1}) s$$

$$= 0$$

$$\therefore \underbrace{E(s|d)}_{\text{solves the above score eq}} = (N^{-1} + \Sigma^{-1})^{-1} N^{-1} d$$

$$= \left(N^{-1}(\Sigma + N)\Sigma^{-1}\right)^{-1} N^{-1} d$$

$$\therefore \boxed{E(s|d) = \underbrace{\Sigma}_{\text{cross cov of } d \text{ with } s.} \underbrace{(\Sigma + N)^{-1}}_{\text{cov matrix of } d} d}$$

Also $-\nabla_s^2 \log p(s,d)$

$$= N^{-1} + \Sigma^{-1}$$

$$\therefore \boxed{var(s|d) = (N^{-1} + \Sigma^{-1})^{-1}}$$

Note: d represents the pixel space data so Σ represents the spatial cov accross pixels.

Computing $(N^{-1} + \Sigma^{-1})^{-1}$ or $(\Sigma + N)^{-1}$ is hard. However if both N & Σ have the same eigen vectors then we get a simplification:

$$N = U \Lambda_n U^T$$

$$\Sigma = U \Lambda_s U^T$$

diag matrix with pos. entries

$$\therefore E(s|d) = U \left(\frac{\Lambda_s}{\Lambda_s + \Lambda_n}\right) U^T d$$

$$var(s|d) = U \left(\frac{1}{\Lambda_s^{-1} + \Lambda_n^{-1}}\right) U^T$$

Take the eigen transform of s, n, d:

$$\tilde{s} := U^T s \sim \mathcal{N}(0, \Lambda_s)$$

$$\tilde{n} := U^T n \sim \mathcal{N}(0, \Lambda_n)$$

$$\tilde{d} := U^T d \sim \mathcal{N}(0, \Lambda_s + \Lambda_d)$$

So $\tilde{d} = \tilde{s} + \tilde{n}$ ← Analogy to the harmonic domain model

where

$$\boxed{E(\tilde{s}|\tilde{d}) = \underbrace{\frac{\Lambda_s}{\Lambda_s + \Lambda_n}}_{\text{every thing is done coordinate wise.}} \tilde{d}}$$

$$\boxed{var(\tilde{s}|\tilde{d}) = \frac{1}{\Lambda_s^{-1} + \Lambda_n^{-1}}}$$

Think of the diag of Λ_s & Λ_n as spectral densities:

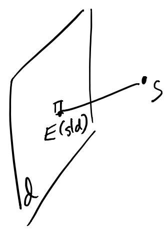
$$E(\tilde{s}_i \tilde{s}_j) = \delta_{ij} (\Lambda_s)_{ii}$$

$$E(\tilde{n}_i \tilde{n}_j) = \delta_{ij} (\Lambda_n)_{ij}$$

Remark:

$E(s|d)$ acts like a projection of d onto s so that

$$s - E(s|d) \perp d$$



where orthogonality in this case is measured by independence (i.e. zero cross cov)

$\therefore$ $var(s - E(s|d))$

$= var(s - E(s|d) \,|\, d)$ ← this is constant when conditioning on d

(✩) $= var(s|d)$

Now to produce a conditional simulation $s^* \sim p(s|d)$ one would normally simulate $z^* \sim N(0, var(s|d))$ & set

$$s^* = E(s|d) + z^*$$

independent

Now if you've coded up $E(s|d)$ & want to avoid computing $var(s|d)$ you can instead produce $\underline{s^* \sim p(s|d)}$ as follows:

step 1: simulate a new data set

$$d^o = s^o + n^o, \quad \begin{cases} s^o \sim N(0, \Sigma) \\ n^o \sim N(0, N) \end{cases}$$

step 2: compute the prediction error $z^o := s^o - E(s^o|d^o)$

step 3: set $s^* = E(s|d) + z^o$ ← indep of $E(s|d)$ & has the right marginal behavior by (✩)

Remark:

If $\tilde{s}^* \sim p(\tilde{s}|\tilde{d})$ then $\tilde{s}^*$ has the right spectral density marginalizing over $\tilde{d}$

$$E(\tilde{s}^*_i \tilde{s}^*_j | \tilde{d}) = E\Big((E(\tilde{s}|\tilde{d}) + \tilde{z}^*)_i (E(\tilde{s}|\tilde{d}) + \tilde{z}^*)_j \,\Big|\, \tilde{d} \Big)$$

$$= E(\tilde{s}|\tilde{d})_i E(\tilde{s}|\tilde{d})_j + E(\tilde{z}^*_i \tilde{z}^*_j)$$

$$= \left(\frac{\Lambda_s}{\Lambda_s + \Lambda_n}\right)_i \left(\frac{\Lambda_s}{\Lambda_s + \Lambda_n}\right)_j \tilde{d}_i \tilde{d}_j + \frac{\delta_{ij}}{(\Lambda_s^{-1} + \Lambda_n^{-1})_{ii}}$$

$$\therefore \quad E(\tilde{s}^*_i \tilde{s}^*_j) = \left(\frac{\Lambda_s}{\Lambda_s + \Lambda_n}\right)_{ii} \left(\frac{\Lambda_s}{\Lambda_s + \Lambda_n}\right)_{jj} \delta_{ij} (\Lambda_s + \Lambda_n)_{ii} + \frac{\delta_{ij}}{(\Lambda_s^{-1} + \Lambda_n^{-1})_{ii}}$$

$$= \left(\frac{\Lambda_s^2}{\Lambda_s + \Lambda_n}\right)_{ii} \delta_{ij} + \left(\frac{\Lambda_s \Lambda_n}{\Lambda_s + \Lambda_n}\right)_{ii} \delta_{ij}$$

$$= (\Lambda_s)_{ii} \, \delta_{ij}$$

$\therefore$ if someone looked at $\tilde{s}^*$ but didn't see $\tilde{d}$ they would think $\tilde{s}^*$ was simply generated from $N(0, \Lambda_s)$.

Remark: Interpretation:

$$E(\tilde{s}|\tilde{d}) = \frac{\Lambda_s}{\underbrace{\Lambda_s + \Lambda_n}} \tilde{d}$$

Represents the proportion of signal power in $\tilde{s}_i$ compared to total power in $\tilde{d}_i$

$$\therefore \quad \text{estimated } \tilde{s}_i \approx \begin{cases} 0 & \text{if proportion is small} \\ \tilde{d}_i & \text{if proportion is large} \end{cases}$$

$$var(\tilde{s}_i|\tilde{d}) = \frac{1}{(\Lambda_s^{-1} + \Lambda_n^{-1})_{ii}} < \min\big((\Lambda_s)_{ii}, (\Lambda_n)_{ii}\big)$$

$$\& \quad var(\tilde{s}_i|\tilde{d}) = \left(\frac{\Lambda_s \Lambda_n}{\Lambda_s + \Lambda_n}\right)_{ii} = \left(\begin{matrix}\text{Proportion} \\ \text{of signal} \\ \text{power}\end{matrix}\right) \cdot \text{noise power}$$

Wiener Filtering

Wiener filtering is just the analog of what we just did but for Stationary / isotropic Gaussian Random fields. ← so the functional eigenvalues of the cov is given by Fourier or spherical Harmonics

Claim (Wiener filter in flat sky):

Suppose $s(x)$ & $n(x)$ are stationary, mean zero, Gaussian Random fields on $x\in\mathbb{R}^d$ (possibly generalized) with spectral densities C_k^{ss}, C_k^{nn} (on $k\in\mathbb{R}^d$) respectively.

Let $d(x) = s(x)+n(x)$.

Then $\hat{s} := E(s|d)$ satisfies
the whole functions

$$\hat{s}_k = \frac{C_k^{ss}}{C_k^{ss}+C_k^{nn}}\, d_k$$

Claim (Wiener filtering on S^2):

Same as above but now

$$d(\hat{n}) = s(\hat{n}) + n(\hat{n})$$

isotropic, mean zero, Gaussian RFs on S^2 with spherical spectral densities C_ℓ^{ss} C_ℓ^{nn}

Then $\hat{s} := E(s|d)$ satisfies

$$\hat{s}_{\ell m} = \frac{C_\ell^{ss}}{C_\ell^{ss}+C_\ell^{nn}}\, d_{\ell m}.$$

Remark: In same sense, Wiener filtering is just Gaussian conditional expected value for the special case of stationary/Isotropic models when $d = s+n$.

Another way to put it:

$$E(Y|X) = \Sigma_{YX}\Sigma_{XX}^{-1} X$$

= "generalized Wiener filtering".

Remark: To produce a conditional sample from $s^* \sim P(s|d)$

on S^2: $$s^*_{\ell m} = \frac{C_\ell^{ss}}{C_\ell^{ss}+C_\ell^{nn}}\, d_{\ell m} + \frac{C_\ell^{ss} C_\ell^{nn}}{C_\ell^{ss}+C_\ell^{nn}}\, \varepsilon_{\ell m}$$

on $\mathbb{R}^d$: $$s^*_k = \frac{C_k^{ss}}{C_k^{ss}+C_k^{nn}}\, d_k + \frac{C_k^{ss} C_k^{nn}}{C_k^{ss}+C_k^{nn}}\, \varepsilon_k$$

When $\varepsilon(\hat{n})$ & $\varepsilon(x)$ are unit var white noise.

Remark: Although $s^* \sim p(s|d)$ behave like marginal draws from $p(s)$, the Wiener filters $\hat{s}(x)$ look different as compared to realizations of $s(x)$.