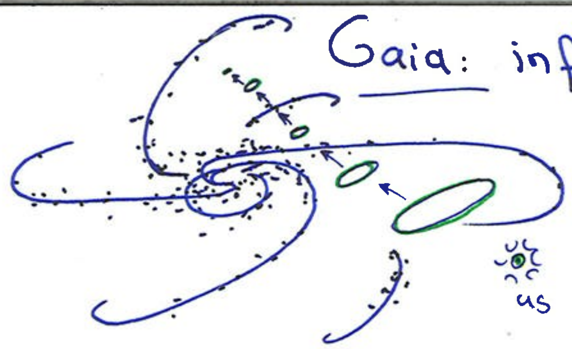
Gaia: inference on the Milky way



: elliptic motions of static stars (idealization) induced by Earth's proper motion around the (static) sun.

• Want: a 3-d map of the stars in the galaxy, given the parallax measurements of Gaia

⇝ the data: little „ellipses" per star ⇝ estimated parallaxes ω

⇝ Q1: which distribution do my incoming data have?

A1: Since Gaia is top precise, we have indeed one of the rare cases where the data scatter Gaussianly:

$$\mathcal{P}(\omega|\omega^t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{[\omega-\omega^t]^2}{\sigma^2}\right)$$

t like „true"

true parallax

⇝ Q2: where is my δ_D-function? = „where is my physics"?

A2: $r_{true} = \frac{1}{\omega^t}$ ⇝ $\delta_D\left(r_t - \frac{1}{\omega^t}\right)$

true distance

true parallax

⇝ So given a measurement of ω, slotting $\mathcal{P}(\omega|\omega^t)$ and δ_D together results in

$$\mathcal{P}(r|\omega) = \int \mathcal{P}(\omega|\omega^t)\, \delta_D\left(r_t - \frac{1}{\omega^t}\right) d\omega^t$$

hence we margin out the unknown true parallax

we only know the estimated ω

$$= \mathcal{P}\left[\omega \,|\, \omega^t = \frac{1}{r}\right] \left|\frac{d\omega}{dr}\right| = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{\exp\left(-\frac{1}{2}\left[\omega_e - \frac{1}{r}\right]^2\right)}{r^2}$$

⇝ This calculation is formally correct, but leads to negative radii if the scatter in parallax is large and if the estimated parallax is negative. [that is not a mistake of the maths, but an accident caused by the Gaia analysis]

⇝ Can we improve upon the situation by using a prior?

- Let's say we have a prior $\Pi(r)$. Due to the Gaussian being symmetric in the sense of

$$„\mathcal{P}(\omega|\omega^\epsilon) \underset{\uparrow}{"="} \mathcal{P}(\omega|\omega^\epsilon)" \text{ in the sense of } [\omega-\omega^\epsilon]^2 = [\omega^\epsilon-\omega]^2,$$

swapping the ϵ over

(if this argument is confusing, then the BHM with $\mathcal{P}(\omega|\omega^\epsilon)$, $\delta_D(r_\epsilon - \frac{1}{\omega_\epsilon})$, $\Pi(r)$ needs to be solved, resulting in the same thing.)

we then have

$$\boxed{\mathcal{P}(r|\omega) = \mathcal{P}(\omega|r_{true})\,\Pi(r_{true})}$$

- This means we are in a situation of a seperate prior: it contributes independent information, so its shape has a meaning which will directly translate into different physical outcomes.

⇝ Slides: what does the „Milkyway" look like, for different priors.

A BHM for reionization & Model selection

- Slides: • what you see, is a completely idealistic simulation of stars eating „ionized bubbles" into neutral hydrogen.
  - these are hypothetical radio observations, where especially the resolution is here at a level no radio telescope can achieve, and we also neglect all uv-plane problems in order to have a pedagogical example that fits into a lecture.

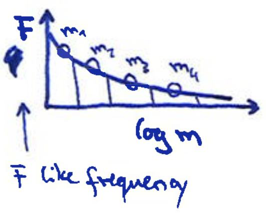
- Physics: the early matter fields in the Universe were low in metals, so cooling via line emission is strongly suppressed when clouds collapse to stars. Result: lacking cooling, a lot of matter is needed to overcome pressure, and the first stars are accordingly very massive. Massive stars are shortlived. („So you can't just hope to observe them „today") They do however power out a lot of ionizing radiation which leads to ionization bubbles in their surrounding matter. If you observe a field of ionization bubbles in the radio [Hydrogen emits at 21cm, which is radio] then you can indirectly infer the properties of the stars that lied in the hydrogen gas.

- Problem: also accretion on early black holes leads to ionizing radiation. The question is therefore not only „what were the properties of my stars" but also „were there already black holes or not, and if yes, how many?"

→ So you have a combined problem: 1) a model selection problem (black holes yes or no?)
2) plus a parameter inference problem (properties of stars + how many black holes if any?)

- Situation 1: no Blackholes ⇝ infer stellar population parameters

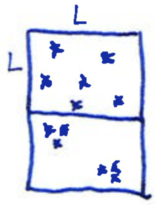
* Stars: mass distribution, e.g.



⇒ $\{m_i\}$ as parameter

* spatial closeness of stars: * d for stellar density in space

* c for clustering behaviour



$d = \frac{6}{L^2}$, $c=0$

$d = \frac{6}{L^2}$, $c=1$

* spectral types of stars (determines output of ionization radiation) "S_1, S_2, S_3"

⇝ Combine all to a parameter vector $\vec{\Theta} = (\{m_i\}, d, c, \{S_i\})$

⇝ let there be a function $f(\vec{\Theta})$ which predicts the power spectrum $C_\ell(\vec{\Theta})$

⇝ in research, this function could be multiple 10.000 lines of radiative transfer code

$$\boxed{\Rightarrow \text{Physics: } \delta_D\left(C_\ell(\vec{\Theta}) - f(\vec{\Theta})\right)}$$

* Data: · a random field, which realizes one single power spectrum: $\hat{C}_\ell \sim \chi^2(\hat{C}_\ell \,|\, C_\ell(\vec{\Theta}))$

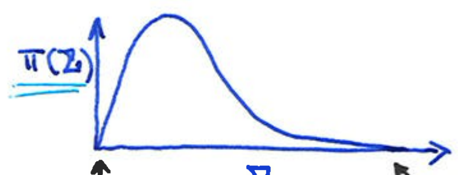
Wishart distribution in real life, here use its univariate analogue: χ^2

on the sky (noisy)

theory (not noisy)

· sadly the field is buried under independent Gaussian shotnoise, of which we only know the upper limit to its variance:

$$n_\ell \sim G(0, \Sigma) \Rightarrow \mathcal{P}(\hat{C}_\ell^n \,|\, \hat{C}_\ell) = G(\mu = \hat{C}_\ell, x = \hat{C}_\ell^n, \Sigma = \Sigma)$$



Gamma-distribution (which probably depends on observational integration time)

probability to get extreme shotnoise supressed

probability that we have zero shotnoise is zero

• Want: $\mathcal{P}(\vec{\theta} \mid \hat{C}^n_\ell)$, $\forall \ell$ the same calculation, hence I drop index ℓ now

$$\rightsquigarrow \mathcal{P}[\vec{\theta} \mid \hat{C}^n] = \int \mathcal{P}(\overset{\text{theory}}{\vec{\theta}}, \overset{\text{random field}}{C}, \hat{C}, \overset{\text{variance of noise}}{\Sigma} \mid \overset{\text{data on the sky}}{\hat{C}^n}) \, dC \, d\Sigma \, d\hat{C}$$

$$= \int \frac{\mathcal{P}(\vec{\theta}, C, \hat{C}, \Sigma, \hat{C}^n)}{\pi(\hat{C}^n)} \, dC \, d\Sigma \, d\hat{C}$$

$$\mathcal{P}(C \mid \vec{\theta}) \, \pi(\vec{\theta}) = \delta_D[C - \overset{\text{code}}{f}(\vec{\theta})] \, \pi(\vec{\theta})$$

Split for physics

$$= \int \frac{\mathcal{P}(\hat{C}, \Sigma, \hat{C}^n \mid \vec{\theta}, C) \, \overbrace{\mathcal{P}(\vec{\theta}, C)}}{\pi(\hat{C}^n)} \, dC \, d\Sigma \, d\hat{C}$$

δ_D removes one integral

noise does not know there are humans with theories in this Universe

$$= \int \frac{\mathcal{P}(\hat{C}, \Sigma, \hat{C}^n \mid \cancel{\vec{\theta}}, f(\vec{\theta})) \, \pi(\vec{\theta})}{\pi(\hat{C}^n)} \, d\Sigma \, d\hat{C}$$

(no dC anymore)

clean up

$$= \int \frac{\mathcal{P}(\hat{C}, \Sigma, \hat{C}^n, f(\vec{\theta})) \, \pi(\vec{\theta})}{\pi(\hat{C}^n) \, \pi[f(\vec{\theta})]} \, d\Sigma \, d\hat{C}$$

Split for the random field

$$= \int \frac{\mathcal{P}(\Sigma, \hat{C}^n \mid \hat{C}, \cancel{f(\vec{\theta})}) \, \overbrace{\mathcal{P}(\hat{C}, f(\vec{\theta}))}^{\mathcal{P}(\hat{C} \mid f(\vec{\theta})) \, \cancel{\pi[f(\vec{\theta})]}} \, \pi(\vec{\theta})}{\pi(\hat{C}^n) \, \cancel{\pi[f(\vec{\theta})]}} \, d\Sigma \, d\hat{C}$$

clean up

$$= \int \frac{\mathcal{P}(\Sigma, \hat{C}^n \mid \hat{C}) \, \pi(\vec{\theta}) \, \mathcal{P}(\hat{C} \mid f(\vec{\theta}))}{\pi(\hat{C}^n)} \, d\Sigma \, d\hat{C}$$

more cleaning for intended noise split

$$= \int \frac{\mathcal{P}(\Sigma, \hat{C}^n, \hat{C}) \, \pi(\vec{\theta}) \, \mathcal{P}(\hat{C} \mid f(\vec{\theta}))}{\pi(\hat{C}^n) \, \pi(\hat{C})} \, d\Sigma \, d\hat{C}$$

Σ and $\hat{C}$ independent (Σ: telescope, $\hat{C}$: physics/universe)

noise split

$$= \int \frac{\mathcal{P}(\hat{C}^n \mid \Sigma, \hat{C}) \, \mathcal{P}(\Sigma, \cancel{\hat{C}}) \, \pi(\vec{\theta}) \, \mathcal{P}(\hat{C} \mid f(\vec{\theta}))}{\pi(\hat{C}^n) \, \cancel{\pi(\hat{C})}} \, d\Sigma \, d\hat{C}$$

$$= \int \frac{\overset{G}{\mathcal{P}(\hat{C}^n \mid \Sigma, \hat{C})} \, \overset{\text{Gamma}}{\mathcal{P}(\Sigma)} \, \overset{\chi^2}{\mathcal{P}(\hat{C} \mid f(\vec{\theta}))} \, \overset{\text{for example flat}}{\pi(\vec{\theta})}}{\pi(\hat{C}^n)} \, d\Sigma \, d\hat{C}$$

$$= \int \frac{G(\hat{C}^n \mid \Sigma, \hat{C}) \, \Gamma(\Sigma) \, \chi^2(\hat{C} \mid f(\vec{\theta})) \, \mathrm{Unif}(\vec{\theta})}{\pi(\hat{C}^n)} \, \underbrace{d\Sigma \, d\hat{C}}$$

we don't know which realization of shotnoise and $\hat{C}$ the Universe has, hence we integrate it out, conditional on our data

What about blackholes now?

• Had (in no-blackholes case):

unaffected $\qquad$ changes $\qquad \mathcal{N}$ $\qquad$ changes $\qquad \mathcal{N}$

$$\mathcal{P}[\vec{\theta} \mid \hat{C}^n] = \int \frac{G(\hat{C}^n \mid \Sigma, \hat{C})\, \Gamma(\Sigma)\, \chi^2(\hat{C} \mid f(\vec{\theta}))\, \mathrm{Unif}(\vec{\theta})}{\Pi(\hat{C}^n)}\, d\Sigma\, d\hat{C}$$

add extra parameter:
$\mathcal{N}$ = number of black holes.
If $\mathcal{N} = 0$ we fall back on the old case
⇝ this is called „a nested model"

what is this? „The probability to get the data at all?"
⇝ this <u>will</u> change: a completely wrong model has zero chance to produce anything like our observed data.

⇝ Adding new parameters to the old ones will either leave the quality of the fit unchanged, but it is probable to improve it

⇝ Adding new parameters also allows more wiggle room [the dimension of param space increases]

$\Rightarrow$ the quality of fit better increase so drastically that the new wiggle room is justified (otherwise what keeps me from adding ever more params?.)

This is what the „evidence" does for you : $\Pi(\hat{C}^n)$