A brief introduction to Bayesian Statistics through Astronomical Applications (Lecture 3)

Cecilia Mateu J.
Centro de Investigaciones de Astronomía (CIDA)

Universidad de Barcelona
26 de octubre de 2016

More Examples

Fitting a density profile to data

- Another example. Lets say you have a catalogue of a particular stellar tracer, e.g. RR Lyrae stars, Red Clump stars, Cepheids or whatever.

- We'd like to use these tracers to study how stars are distributed in space in the different components in the Galaxy, i.e. their density profile ρ

- How do we compare and fit this function to a sample of stars?

Thick Disc and Halo RRL Density Profiles

- Thick Disc density profile

$$\rho_{tkd} = e^{-\frac{R-R_\odot}{h_R}} e^{-\frac{|z|}{h_z}}$$

- Halo density profile

$$\rho_H = \frac{1}{R_\odot^n}\left[R^2 + z^2\right]^{n/2}$$

- Total density profile:

$$\rho = N_{mod}\left(\rho_{tkd} + C_{hd}\rho_H\right) = N_{mod}\hat{\rho}$$

$$N_{RR} = \iiint_{V_S} \rho(\vec{r})dV$$

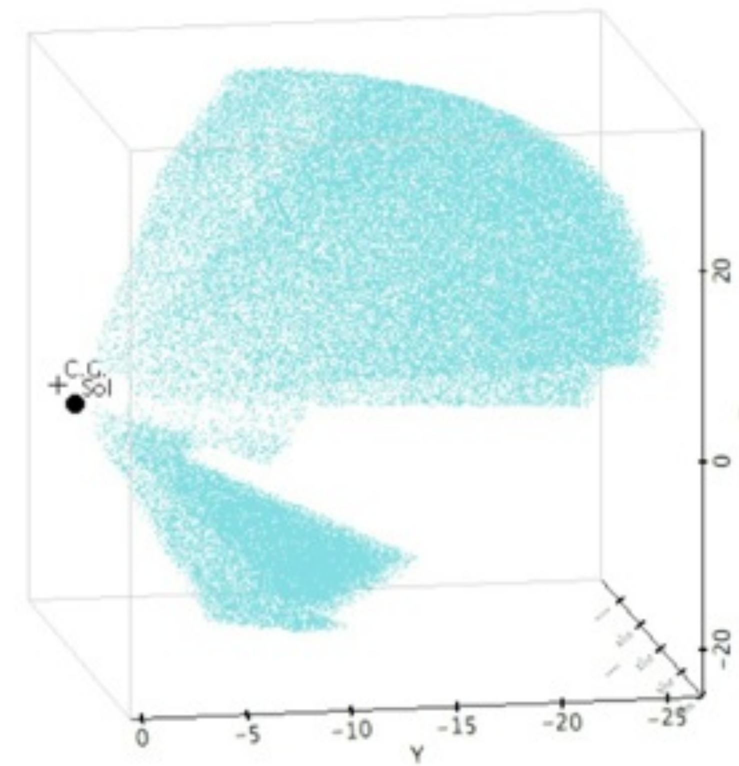
... but the density ρ is not a directly observable quantity! all we have are the total number N_{OBS} of RRLs in the survey volume V_S and their observed positions

Density Profiles: A Bayesian approach

- Our free parameters are:

$$\vec{\theta} = \{(h_z, h_R, n, \hat{C}_{hd}), N_{mod}\}$$

- Lets write our likelihood function as the probability that our model predicts a total number N_{mod} of RRLs in the survey volume, times the probability that each RRL was observed at a position $\vec{r}_i$



$$L(N_{obs}, \{\vec{r}_i\}|\vec{\theta}) = \underbrace{P(N_{obs}|\vec{\theta})}_{e^{-N_{mod}}(N_{mod})^{N_{obs}}} \prod_{i=1}^{N_{obs}} \underbrace{P(\{\vec{r}_i\}|\vec{\theta})}_{\hat{\rho}(\vec{r}_i|\vec{\theta})}$$

Poisson probability of observing N_{obs} given predicted number N_{mod}

Density profile (normalized to integrate 1 over the survey volume)

Lombardi et al. (2013)

Density Profiles: A Bayesian approach

- So we have

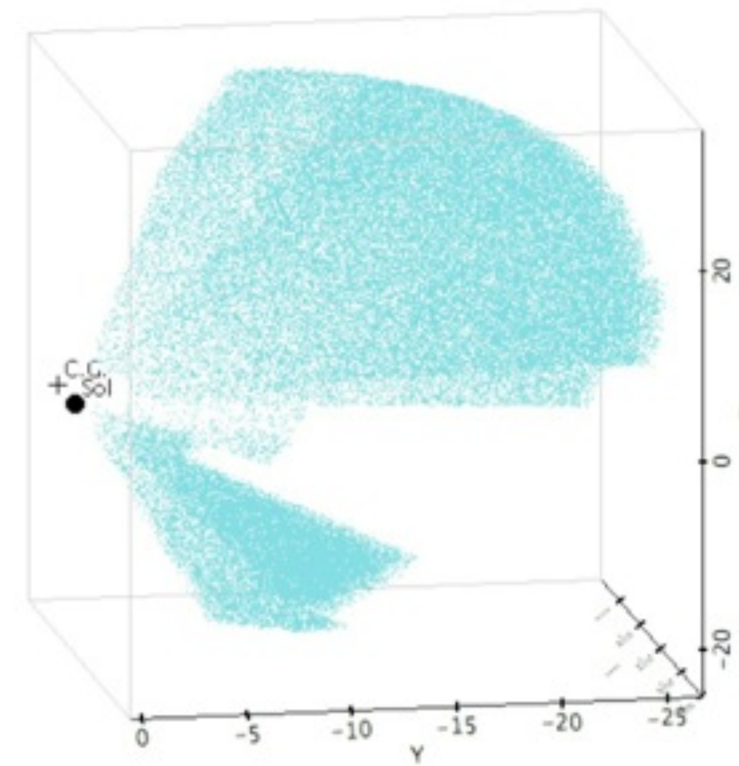
$$\ln L(N_{obs}, \{\vec{r}_i\}|\vec{\theta}) = \left(e^{-N_{mod}}(N_{mod})^{N_{obs}}\right)\left(\prod_{i=1}^{N_{obs}} \hat{\rho}(\vec{r}_i|\vec{\theta})\right)$$

- which we can write as a log-likelihood for a more convenient computation as:

$$\ln L(N_{obs}, \{\vec{r}_i\}|\vec{\theta}) = \left(N_{obs}\ln N_{mod} - N_{mod}\right) + \sum_{i=1}^{N_{obs}} \ln \hat{\rho}(\vec{r}_i|\vec{\theta})$$

Lombardi et al. (2013)

Density Profiles: A Bayesian approach



- Our free parameters are:

$$\vec{\theta} = \{(h_z, h_R, \hat{C}_{tkd}, n, \hat{C}_h), N_{model}^{RRL}\}$$

- Lets write our likelihood function as the probability that our model predicts a total number N_{model}^{RRL} of RRLs in the survey volume, times the probability that each RRL was observed at a position $\vec{r}_i$

$$L(N_{obs}^{RRL}, \{\vec{r}_i\}|\vec{\theta}) = \prod_{i=1}^{N_{obs}^{RRL}} \underbrace{P(N_{RRL}|\vec{\theta})}_{e^{-\mu}\mu^{N_{obs}^{RRL}}} \underbrace{P(\{\vec{r}_i\}|\vec{\theta})}_{\hat{\rho}(\vec{r}_i|\vec{\theta})}$$

Poisson probability of observing N_{obs}^{RRL} given predicted number N_{model}^{RRL}

Density profile (normalized to integrate 1 over the survey volume)

Lombardi et al. (2013)

Density Profiles: A Bayesian approach

- So we have

$$\ln L(N_{obs}^{RRL}, \{\vec{r}_i\}|\vec{\theta}) = \left(e^{-N_{model}^{RRL}} {N_{model}^{RRL}}^{N_{obs}^{RRL}}\right)\left(\prod_{i=1}^{N_{obs}^{RRL}} \hat{\rho}(\vec{r}_i|\vec{\theta})\right)$$

- which we can write as a log-likelihood for a more convenient computation as:

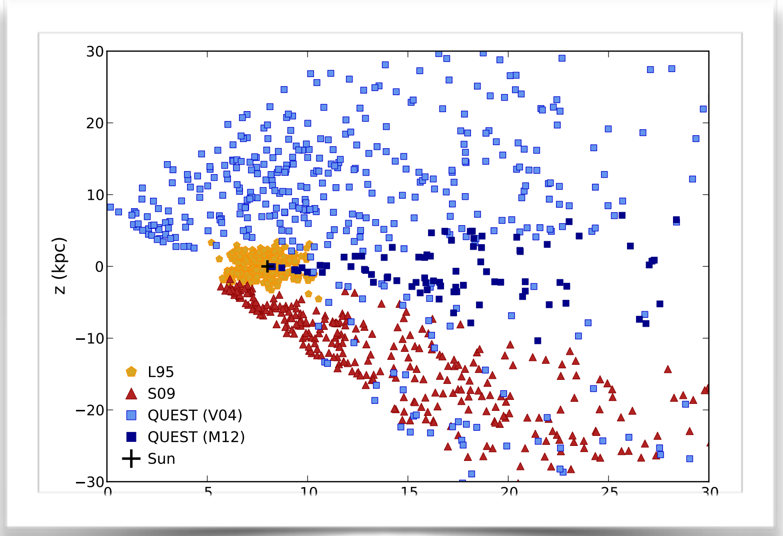
$$\ln L(N_{obs}^{RRL}, \{\vec{r}_i\}|\vec{\theta}) = \left(N_{obs}^{RRL} \ln N_{model}^{RRL} - N_{model}^{RRL}\right) + \sum_{i=1}^{N_{obs}^{RRL}} \ln \hat{\rho}(\vec{r}_i|\vec{\theta})$$

This is the computationally intensive part

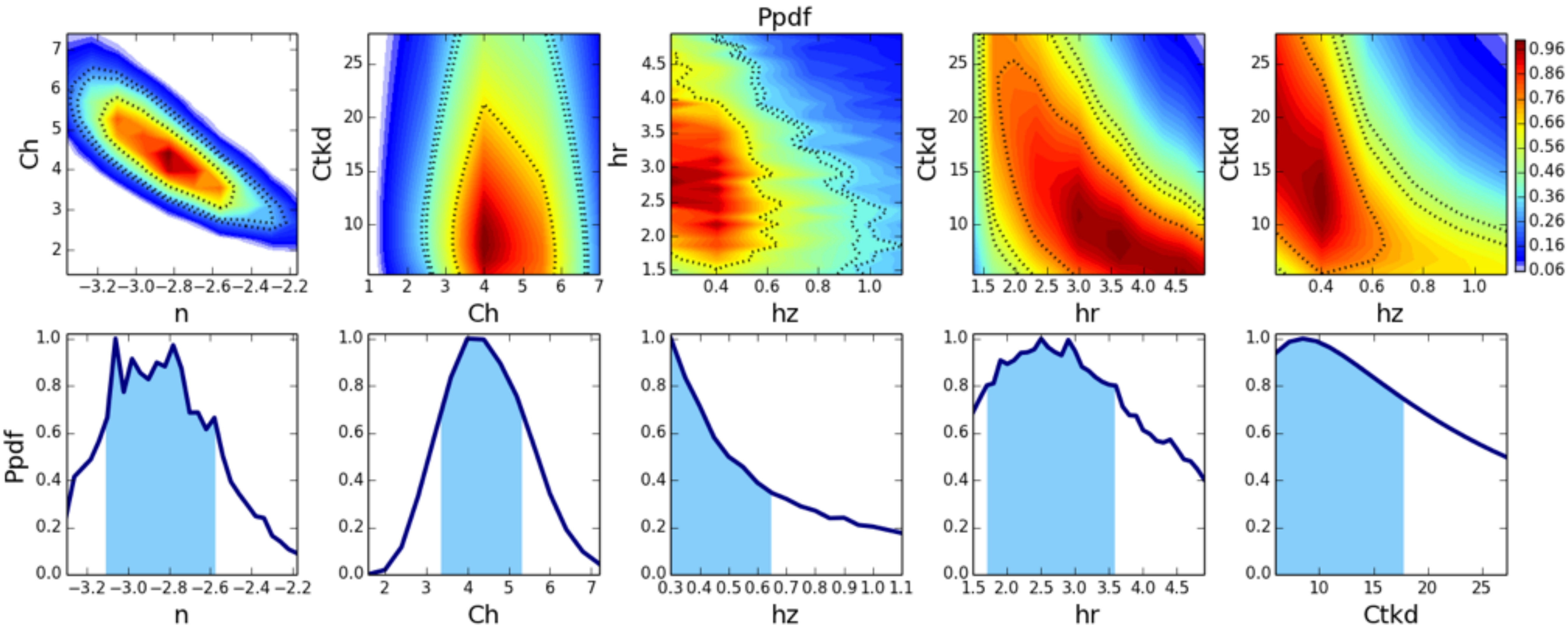
$$N_{RR} = \iiint_{V_S} \rho(\vec{r})dV$$

Lombardi et al. (2013)

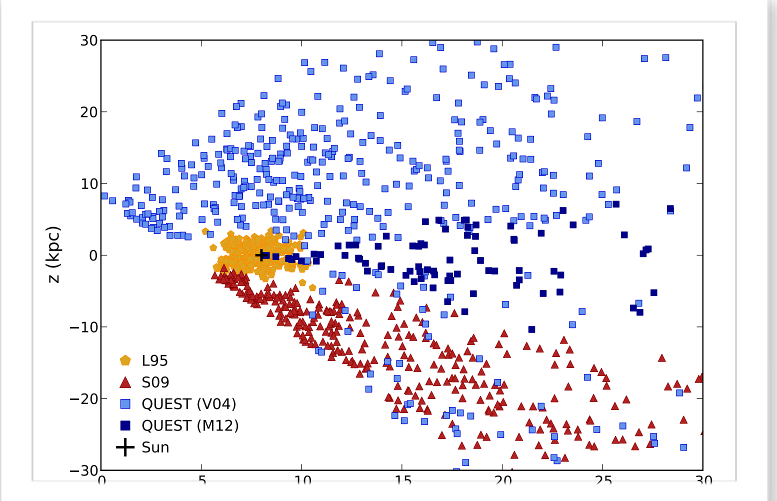
Density Profiles: One sample at a time



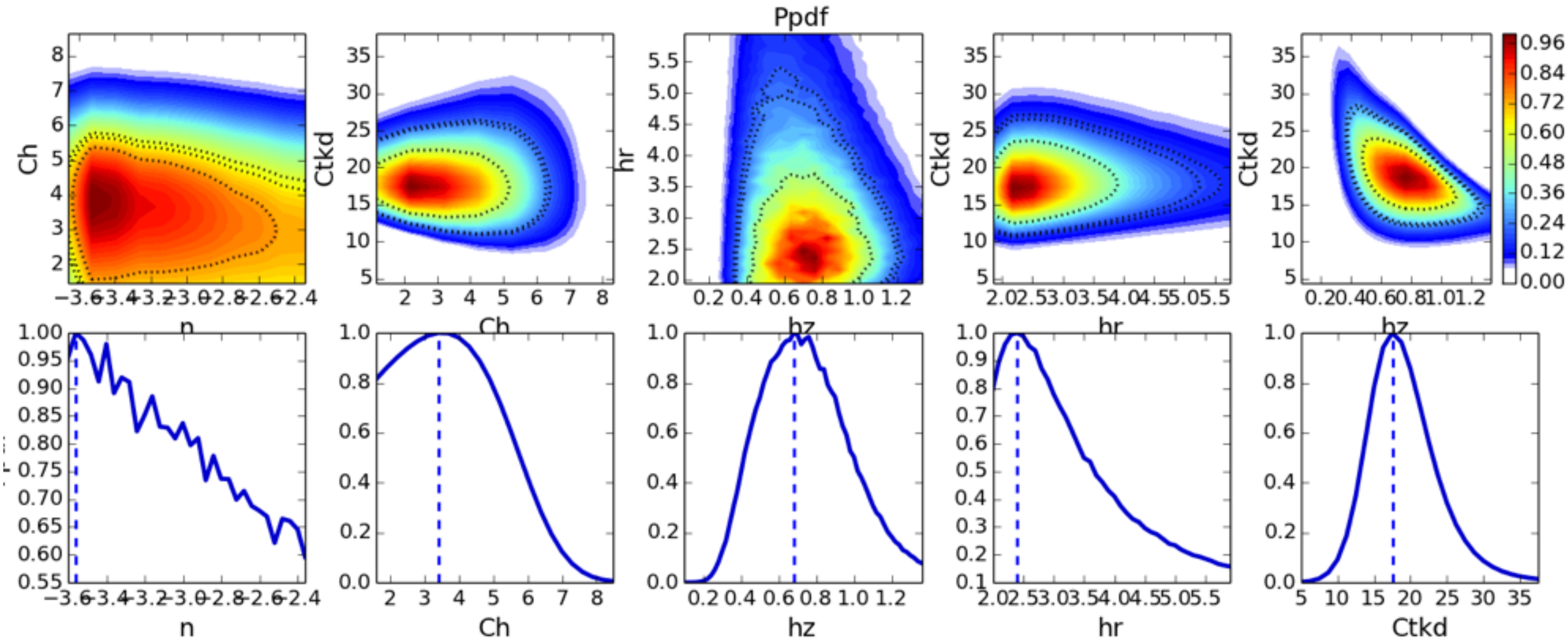
The marginal posteriors taking only the **QUEST sample** are as follows:



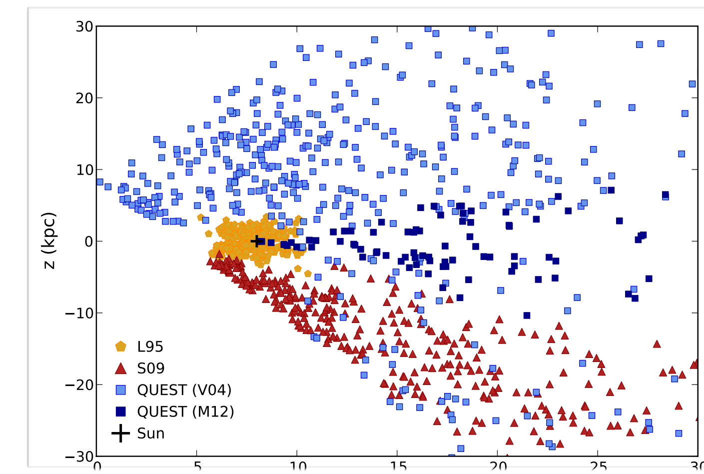
Density Profiles: One sample at a time



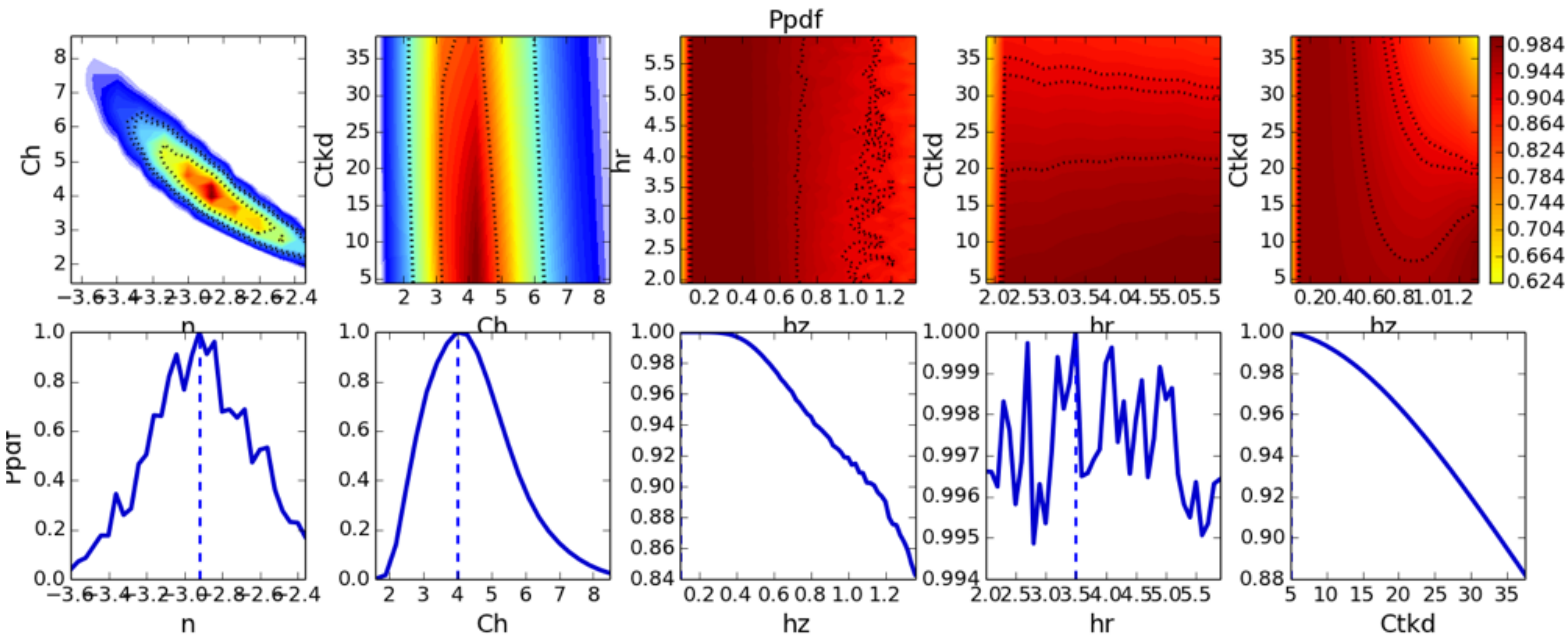
The marginal posteriors taking only the **Layden sample** are as follows:



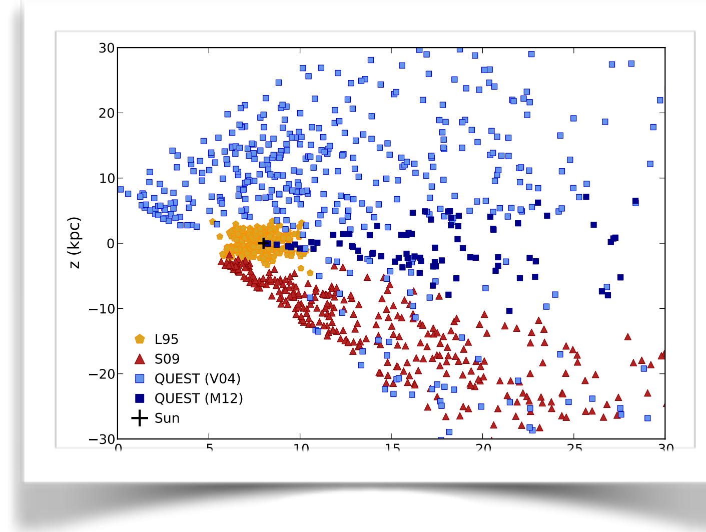
Density Profiles: One sample at a time



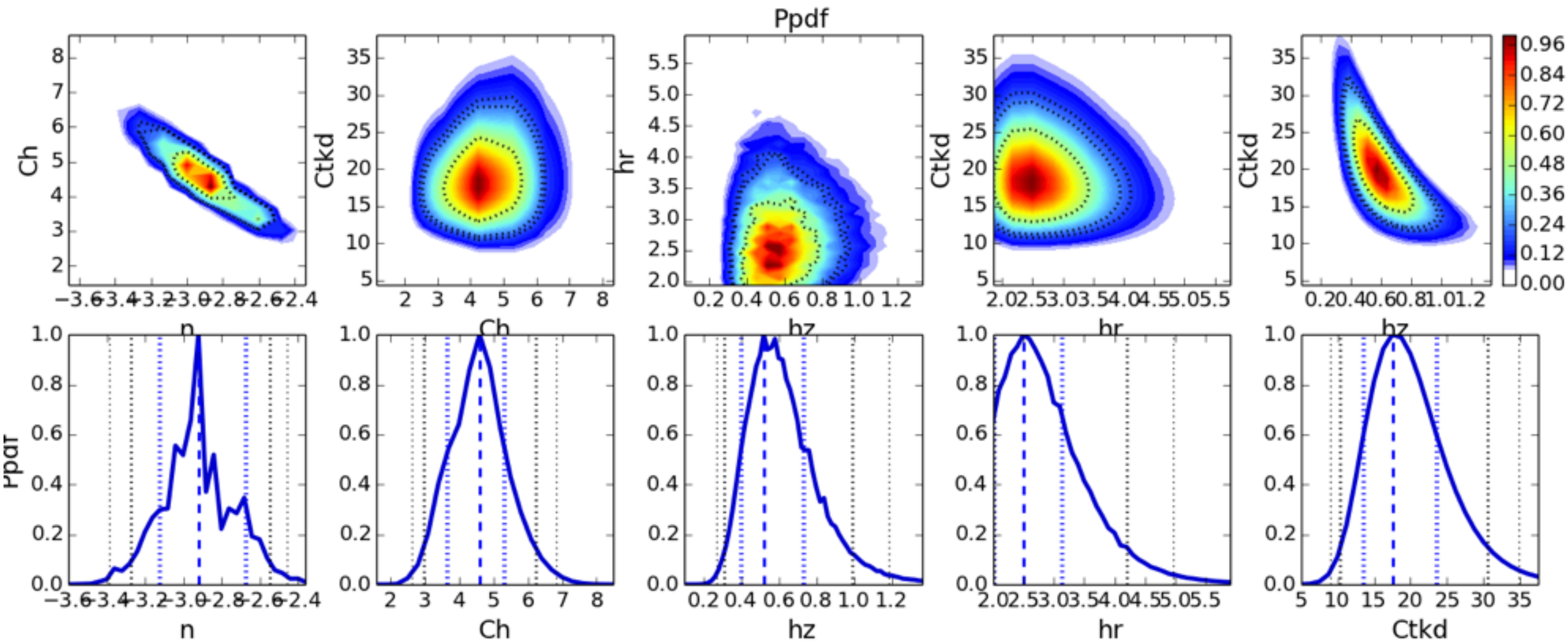
The marginal posteriors taking only the **Sesar sample** are as follows:



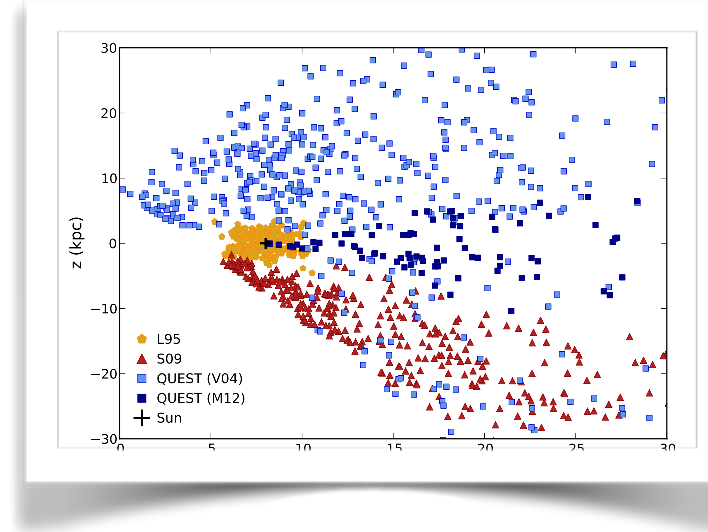
Density Profiles: Combined samples



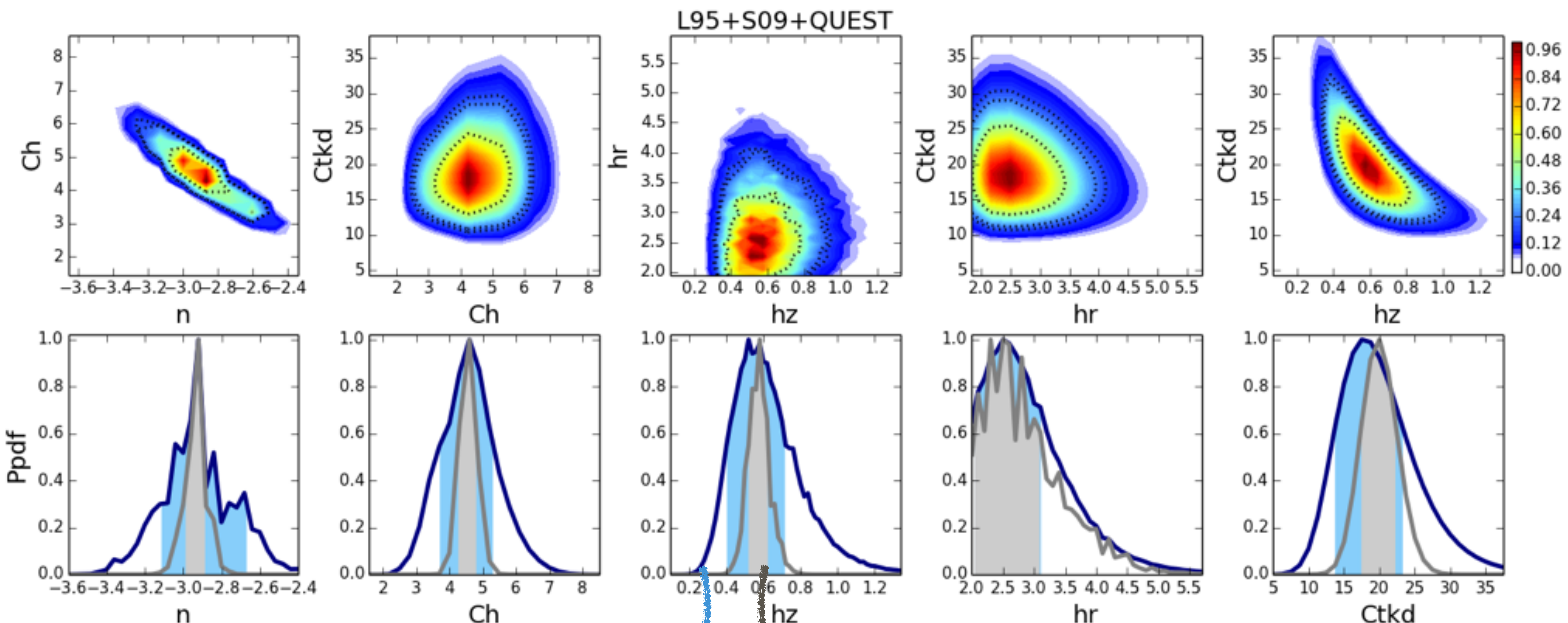
Combining three different samples we find these marginal posteriors (remember its the product of the individual pdfs in 5-D space, and then marginalizing):



Density Profiles: Combined samples



Combining three different samples we find these marginal posteriors (remember its the product of the individual pdfs in 5-D space, and then marginalizing):



marginal distribution

conditional distribution

More Examples

Coordinate Transformations

- Lets say we have a problem in which we know how to write our posterior P(x) in a variable x, but in reality we are interested in knowing the posterior P(y) for variable y=f(x)

- We know the integral of the probability must be conserved

$$\int P(x)dx = \int P(y)dy \qquad\qquad P(x)dx = P(y)dy$$

- This makes it easy, we know how to change variables inside an integral

$$P(y) = \left|\frac{dx}{dy}\right| P(x)$$

Jacobian of the transformation f(x)

The Parallax Example

- Lets illustrate this with an example. Say we have N measurements of the parallax for a star from Gaia
- Lets assume we have an error model and CU7/DPAC gives us an estimate on the parallax error and tells us that its safe to assume these errors are gaussian
- What we'd ultimately like to get is the posterior on the distance to the star

The Parallax Example

- Lets write the posterior on the parallax first

$$P(\varpi|\{\varpi_i\}) = P(\{\varpi_i\}|\varpi)P(\varpi) \qquad |I$$

- Our likelihood is

$$P(\{\varpi_i\}|\varpi) = \prod_{i=1}^{N} e^{-\frac{(\varpi-\varpi_i)^2}{2\sigma_\varpi^2}}$$

- So the posterior is

$$P(\varpi|\{\varpi_i\}) = \prod_{i=1}^{N} e^{-\frac{(\varpi-\varpi_i)^2}{2\sigma_\varpi^2}} P(\varpi) \qquad |I$$

no coordinate transformations yet...

The Parallax Example

- Now we want the posterior for the distance D=1/**ϖ**

$$P(\varpi|\{\varpi_i\})d\varpi = P(D|\{\varpi_i\})dD$$

$$P(D|\{\varpi_i\}) = P(\varpi|\{\varpi_i\})\left|\frac{d\varpi}{dD}\right|$$

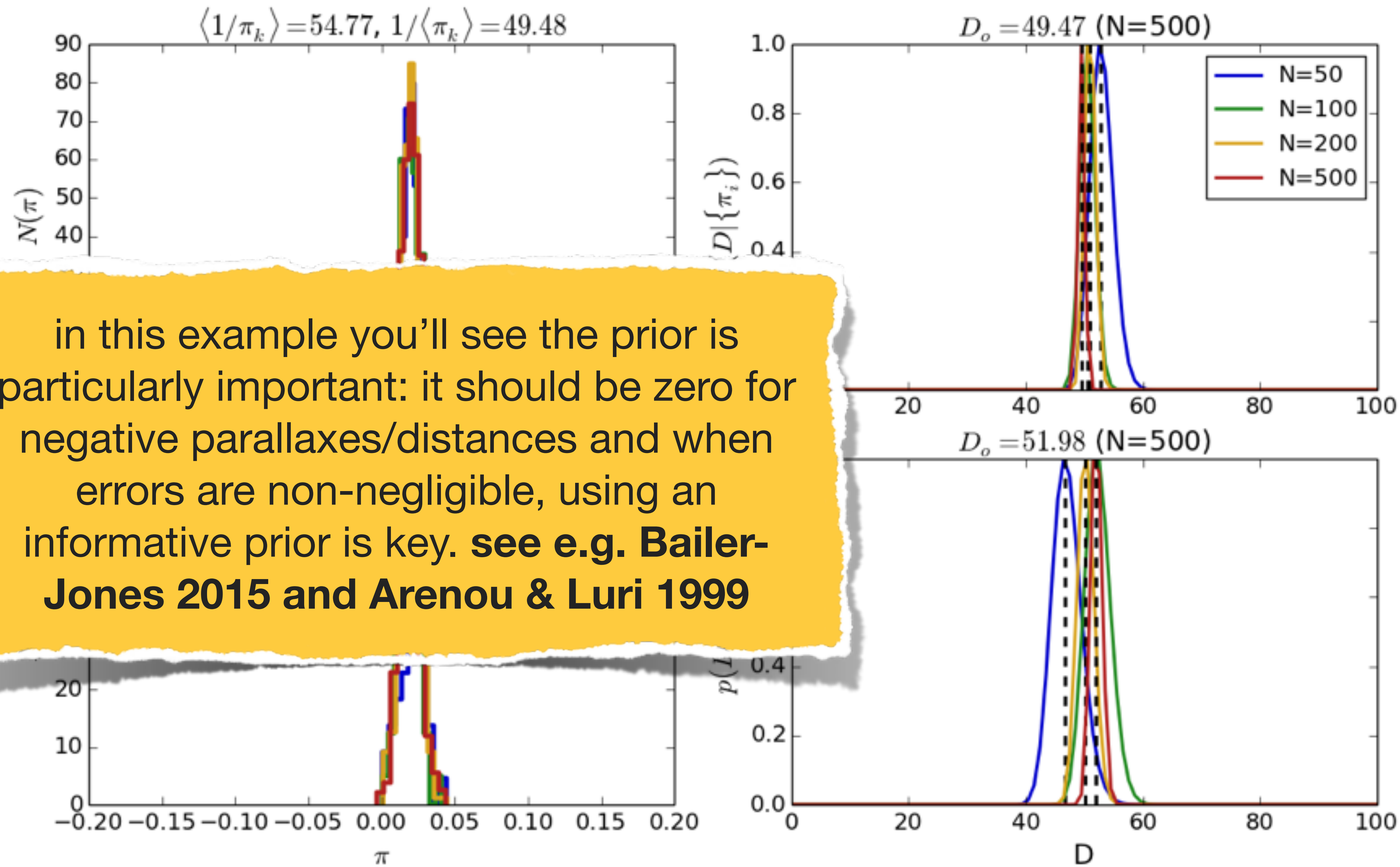
$$P(\varpi|\{\varpi_i\}) = \prod_{i=1}^{N} e^{-\frac{(\varpi-\varpi_i)^2}{2\sigma_\varpi^2}} P(\varpi)$$

$$P(D|\{\varpi_i\}) = P(\varpi|\{\varpi_i\})\frac{1}{D^2}$$

$$P(D|\{\varpi_i\}) = \prod_{i=1}^{N} e^{-\frac{(1/D-\varpi_i)^2}{2\sigma_\varpi^2}} \frac{1}{D^2} P(\varpi(D))$$

The Parallax Example

(go to paralaje.py and play around)



in this example you’ll see the prior is particularly important: it should be zero for negative parallaxes/distances and when errors are non-negligible, using an informative prior is key. **see e.g. Bailer-Jones 2015 and Arenou & Luri 1999**

Error Propagation

Error Propagation

- Lets say we know the PDF for two variables X and Y

$$P(X) \qquad p(Y) \qquad\qquad |data, I$$

- Now we'd like to know what is the PDF for Z, where

$$Z = X + Y$$

- First, lets look at the case when X and Y are conditionally independent. If this is the case, the joint probability P(X,Y)=P(X)p(Y), from the Marginalization rule this is:

$$P(Z) = \iint P(X)p(Y)dXdY \qquad |data, I$$

for all X,Y
t.q. Z=X+y

Error Propagation

- First, lets look at the case when X and Y are conditionally independent. If this is the case, the joint probability P(X,Y)=P(X)p(Y), from the Marginalization rule this is:

$$P(Z) = \iint_{\substack{\text{for all X,Y} \\ \text{t.q. Z=X+Y}}} P(X)p(Y)dXdY \qquad |data, I$$

$$P(Z) = \iint P(X)p(Y)\delta(Z - [X+Y])dXdY$$

$$P(Z) = \int P(X)p(Y = Z - X)dX$$

note this is the convolution of P(X) and p(Y)

Error Propagation

- If P(X) and p(y) are gaussians, e.g. lets say like yesterday we have a model with gaussian uncertainties and a uniform prior,

$$P(X) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{(X-X_o)^2}{2\sigma_X^2}} \qquad P(Y) = \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{(Y-Y_o)^2}{2\sigma_Y^2}}$$

$$|data, I$$

then

$$P(Z) = \int P(X)p(Y = Z - X)dX$$

is the convolution of these two gaussians

$$P(Z) = \frac{1}{2\pi\sigma_X\sigma_Y} \int_{-\infty}^{+\infty} e^{-\frac{(X-X_o)^2}{2\sigma_X^2}} e^{-\frac{(Z-X-Y_o)^2}{2\sigma_Y^2}} dX$$

Error Propagation

- After completing squares and simplifying we get

$$P(Z) = \frac{1}{2\pi\sigma_z} e^{-\frac{(Z-Z_o)^2}{2\sigma_z^2}}$$

where

$$Zo = X_o + Y_o \qquad \text{and} \qquad \sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$$

so, the sum in quadrature rule is derived

- Lets go back to the general problem, where Z=f(X,Y)

Error Propagation

- In general, i.e. without assuming cond. independence of X,Y we have

$$P(Z) = \iint\limits_{\substack{\text{for all X,Y} \\ \text{t.q.} \\ \text{Z=f(X,Y)}}} P(X,Y)dXdY$$

- Since Z=f(X,Y) we have

$$P(Z) = \iint P(X,Y)\delta(Z - f(X,Y))dXdY$$

The Parallax Example

- Now we want the posterior for the distance D=1/$\boldsymbol{\varpi}$

$$P(\varpi|\{\varpi_i\})d\varpi = P(D|\{\varpi_i\})dD$$

$$P(D|\{\varpi_i\}) = P(\varpi|\{\varpi_i\})\left|\frac{d\varpi}{dD}\right|$$

$$P(\varpi|\{\varpi_i\}) = \prod_{i=1}^{N} e^{-\frac{(\varpi-\varpi_i)^2}{2\sigma_\varpi^2}} P(\varpi)$$

$$P(D|\{\varpi_i\}) = P(\varpi|\{\varpi_i\})\frac{1}{D^2}$$

$$P(D|\{\varpi_i\}) = \prod_{i=1}^{N} e^{-\frac{(1/D-\varpi_i)^2}{2\sigma_\varpi^2}} \frac{1}{D^2} P(\varpi(D))$$