An introduction to Bayesian Statistics through Astronomical Applications (Lecture 2)

Cecilia Mateu J.
Centro de Investigaciones de Astronomía (CIDA)
Mérida, Venezuela

Universidad de Barcelona
21 de noviembre de 2017

The Coin Example

Github Repository

* Classes and programs are available in the following GitHub repository

https://github.com/cmateu/intro_to_bayes_UB

The Coin Example

* Lets recap
  * N and N_h are our data (known)
  * Our goal is to get $P(h|N,N_h,I)$ remember this is a function of h

The full posterior IS the answer to our problem

$$P(h|N,N_h,I) = Ch^{N_h}(1-h)^{N-N_h}\,P(h|I)$$

The Coin Example

The full posterior IS the answer to our problem

$$P(h|N,N_h,I) = Ch^{N_h}(1-h)^{N-N_h}\, P(h|I)$$

* anything else we may want can be calculated from it, e.g.
  * the most probable value of h
  * credible regions (Bayesian term for confidence intervals)
  * The probability that $p>0.5$
    * int $P(p|N,n,I)\,dp$
  * ... more on this ...

The Coin Example

The full posterior IS the answer to our problem

$$P(h|N,N_h,I) = Ch^{N_h}(1-h)^{N-N_h}\,P(h|I)$$

* Question:
    * is it equivalent to take the data as a whole or to take a subset and add new data as it comes?

Updating Information

Updating Information

* Lets consider the case of having two independent data points D_1 and D_2. Bayes' Theorem states

$$P(H|D,I) \propto \prod_{i=1,2} P(D_i|H,I)\; P(H|I)$$

* Expanding the product in the likelihood term:

$$P(H|D,I) \propto P(D_2|H,I)\; P(D_1|H,I)\; P(H|I)$$

Updating Information

* Lets consider the case of having two independent data points D_1 and D_2. Bayes' Theorem states

$$P(H|D,I) \propto \prod_{i=1,2} P(D_i|H,I)\, P(H|I)$$

* Expanding the product in the likelihood term:

$$P(H|D,I) \propto P(D_2|H,I)\, \underbrace{P(D_1|H,I)\, P(H|I)}_{P(H|D_1,I)}$$

Updating Information

* Lets consider the case of having two independent data points D_1 and D_2. Bayes' Theorem states

$$P(H|D_1,D_2,I) \propto \prod_{i=1,2} P(D_i|H,I)\ P(H|I)$$

* Expanding the product in the likelihood term:

$$P(H|D_1,D_2,I) \propto P(D_2|H,I)\ \boxed{P(D_1|H,I)\ P(H|I)}$$

$$P(H|D_1,D_2,I) \propto P(D_2|H,I)\,P(H|D_1,I)$$

Updating Information

* Lets consider the case of having two independent data points D_1 and D_2. Bayes' Theorem states

$$P(H|D_1,D_2,I) \propto \prod_{i=1,2} P(D_i|H,I)\; P(H|I)$$

* Expanding the product in the likelihood term:

$$P(H|D_1,D_2,I) \propto P(D_2|H,I)\; P(D_1|H,I)\; P(H|I)$$

$$P(H|D_1,D_2,I) \propto P(D_2|H,I)\, P(H|D_1,I)$$

* Here $P(H|D_1)$ the posterior on H given D_1 is acting as an updated prior!

The Coin Example

The full posterior IS the answer to our problem

$$P(h|N,N_h,I) = Ch^{N_h}(1-h)^{N-N_h}\,P(h|I)$$

* anything else we may want can be calculated from it, e.g.
  * the most probable value of h
  * credible regions (Bayesian term for confidence intervals)
  * The probability that p>0.5
    * int P(p|N,n,I) dp
  * ... more on this ...

More examples

The full posterior IS the answer to our problem

$$P(h|N, N_H, I) \propto h^{N_H}(1-h)^{N-N_H}P(h|I)$$

More examples

The full posterior IS the answer to our problem

$$P(h|N, N_H, I) \propto h^{N_H}(1-h)^{N-N_H} P(h|I)$$

* The probability that the coin is biased:
    * Lets say if 0.45<h<0.55 we can safely take the coin as fair

More examples

The full posterior IS the answer to our problem

$$P(h|N, N_H, I) \propto h^{N_H}(1-h)^{N-N_H} P(h|I)$$

* The probability that the coin is biased:
    * Lets say if 0.45<h<0.55 we can safely take the coin as fair

$$P_{fair} = \int_{0.45}^{0.55} P(h|N, N_H, I) dh$$

More examples

The full posterior IS the answer to our problem

$$P(h|N, N_H, I) \propto h^{N_H}(1-h)^{N-N_H} P(h|I)$$

* The probability that the coin is biased:
    * Lets say if $0.45<h<0.55$ we can safely take the coin as fair

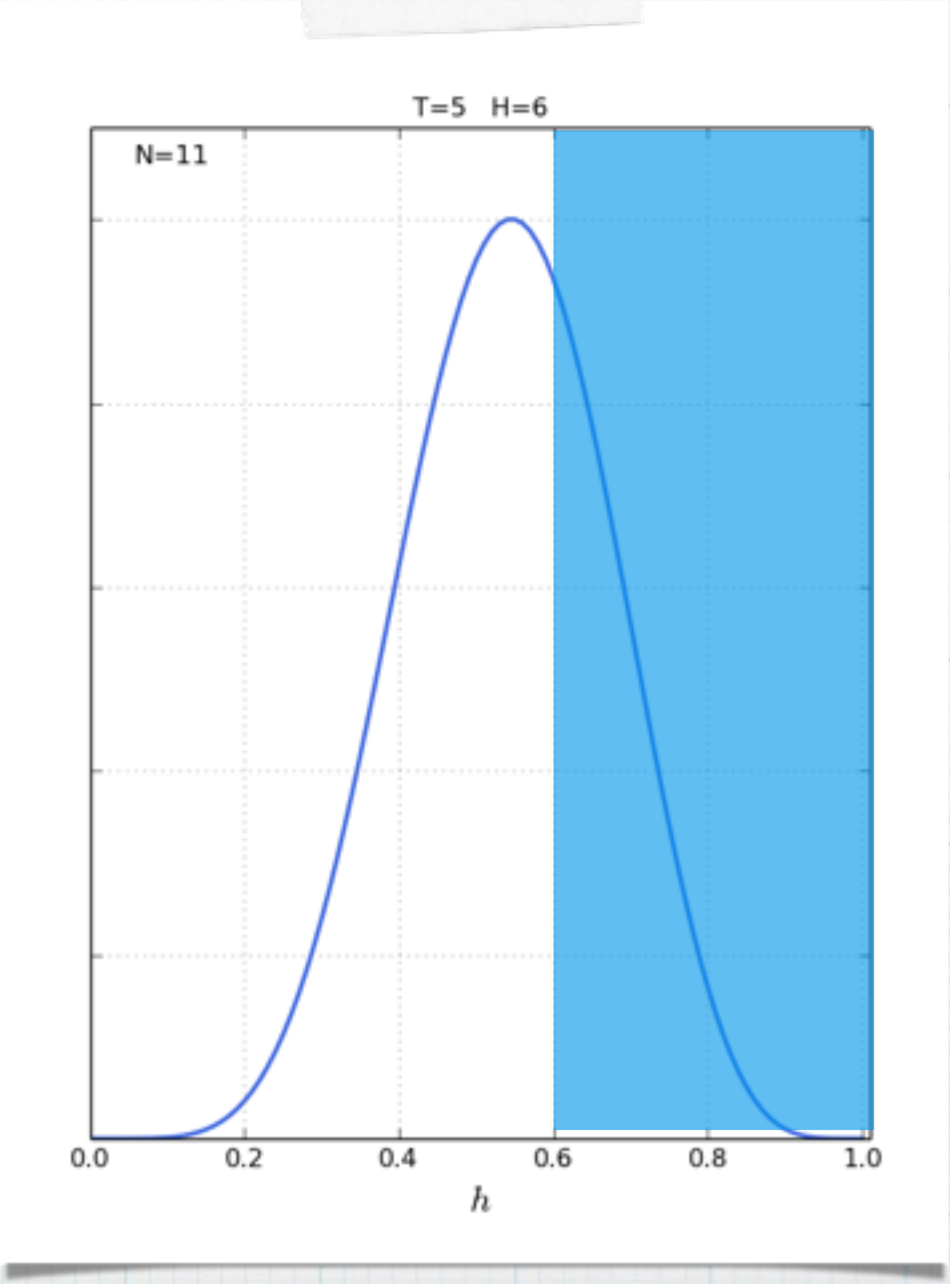
$$P_{fair} = \int_{0.45}^{0.55} P(h|N, N_H, I)dh$$

* so, the probability that it is biased is $P_{biased} = 1 - P_{fair}$

$$P_{biased} = \int_{0}^{0.45} P(h|N, N_H, I)dh + \int_{0.55}^{1.} P(h|N, N_H, I)dh$$

Marginalization

* We want to compute the probability that the coin is biased towards heads
* Lets say by this, we mean h>0.6
* We have to integrate the posterior over the desired range of h



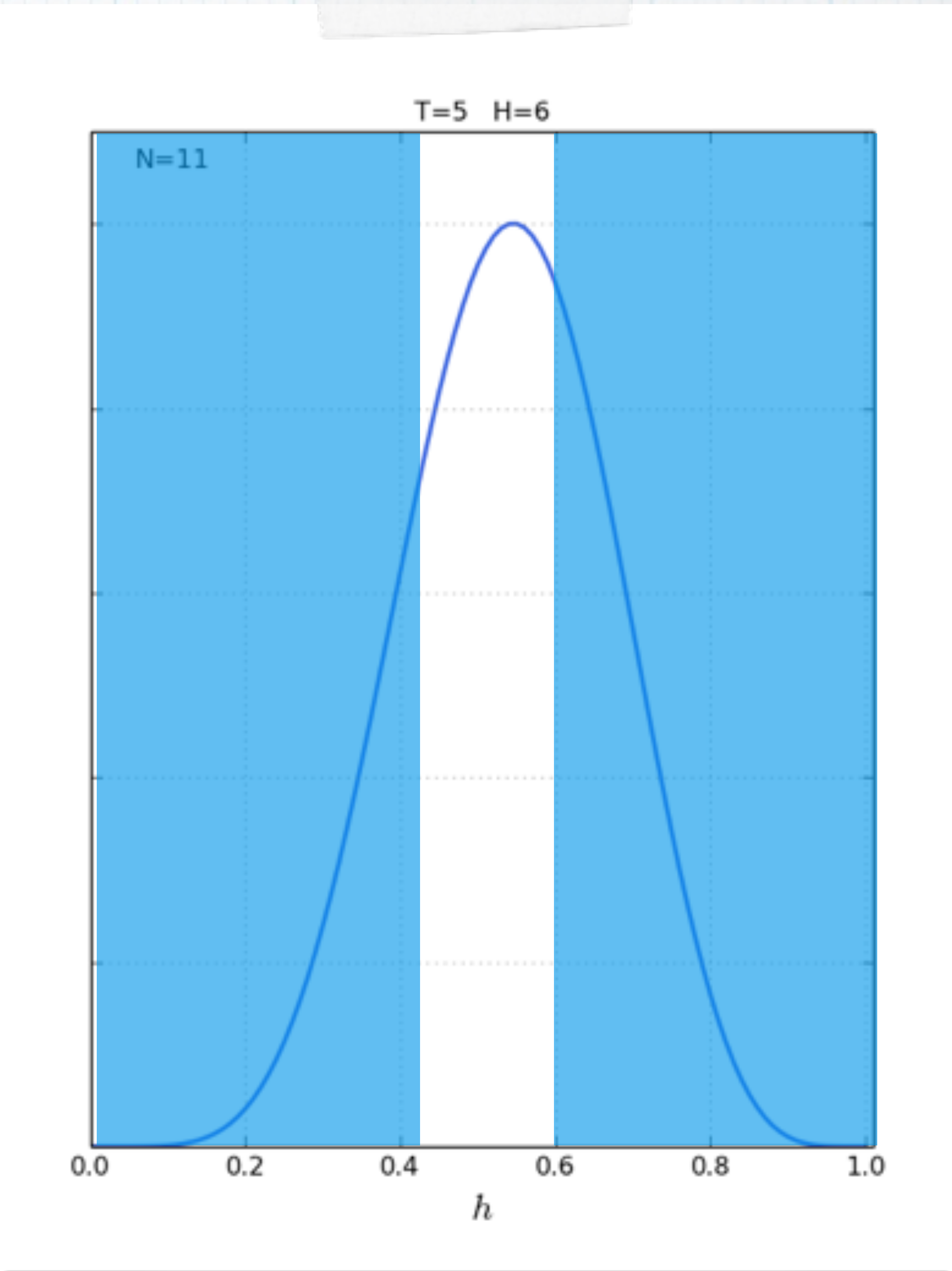
$$P(h > 0.6 | N_h, N) = \int_{0.6}^{1} P(h | N_h, N) dh$$

Marginalization

* Now, lets compute the probability that the coin is fair

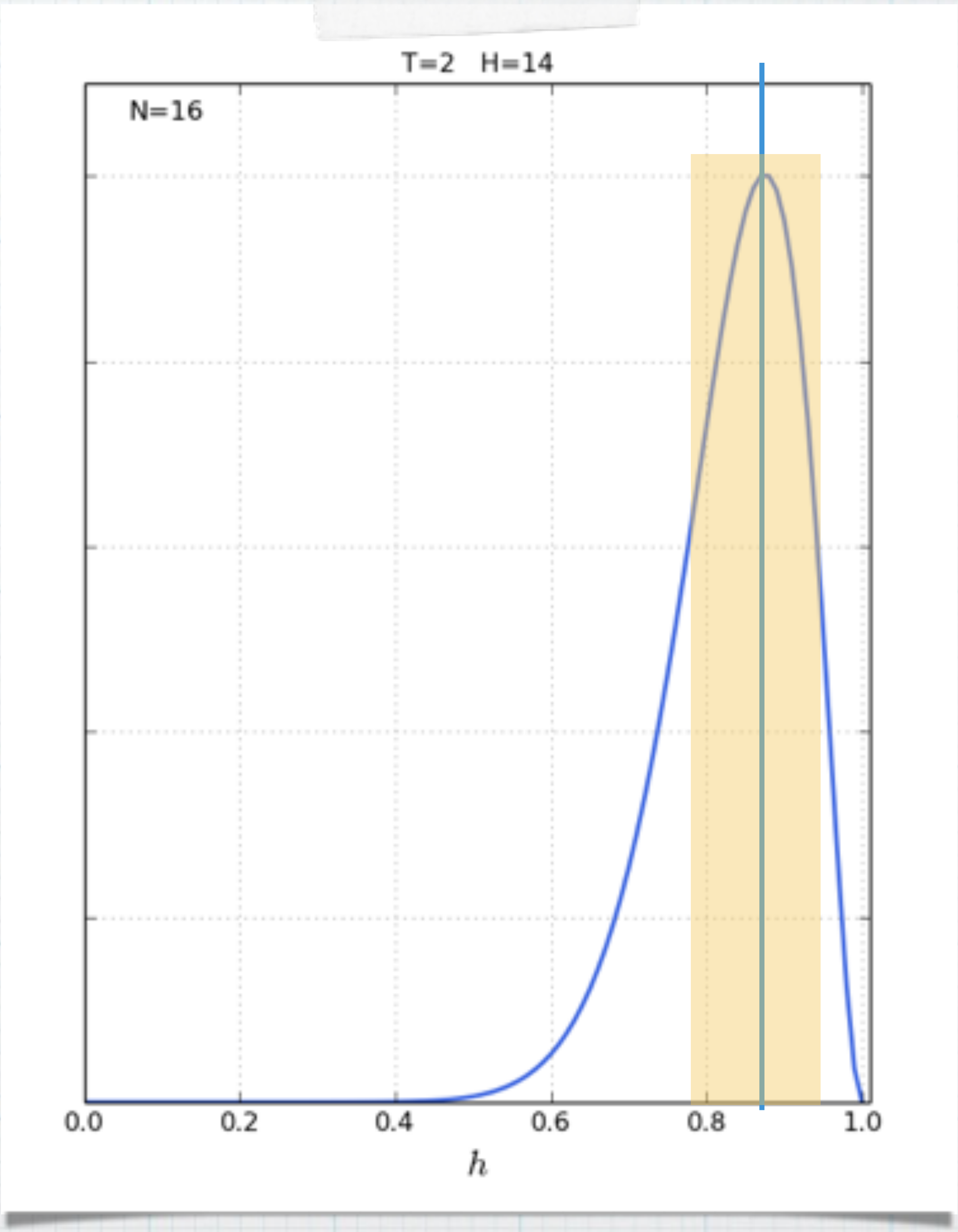
* Lets say by fair we mean h=0.5 ± x, where x could be e.g. x=0.05

$$P(|h-0.5|<x|N_h,N)$$
$$=\int_{0.5-x}^{0.5+x} P(h|N_h,N)dh$$



More Things to compute

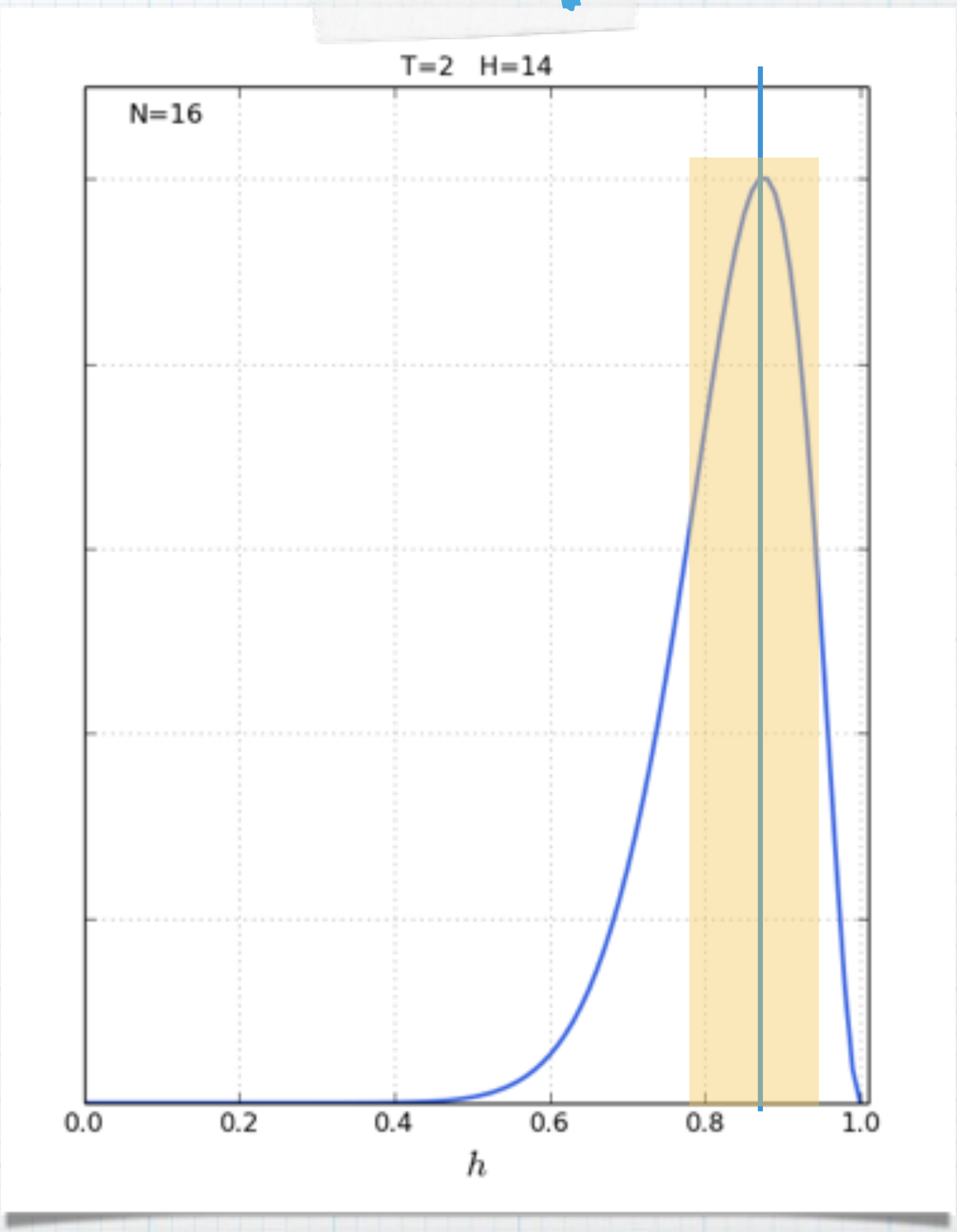
- h_0 = Most probable value of h, i.e. h where $P(h|N_h,N)$ is maximum
- Credible regions:
  - An X% credible region contains X% of the area of the posterior
  - e.g. 1σ intervals are a 68% credible region for a gaussian posterior



More Things to compute

* Lets report as an error bar, the 68% credible region

* The most probable value is Nh/N, the usual answer, but there's a natural way of computing the error bars



* This is specially important for extremely low or extremely high values of h

The Coin Example in an astrophysical context

Disk Fractions

* The Coin is just one example of a Binomial problem
* This describes anything that can be expressed as a two-state problem, a 'success' occurring with probability p and 'failure' with probability (1-p), for example p could be:
    * The fraction of radio-loud quasars in a sample
    * The fraction of stars having disks
    * The fraction of early/late type galaxies
    * ....

Disk Fractions

* The Coin is just one example of a Binomial problem
* This describes anything that can be expressed as a two-state problem, a 'success' occurring with probability p and 'failure' with probability (1-p), for example p could be:
    * The fraction of radio-loud quasars in a sample
    * The fraction of stars having disks
    * The fraction of early/late type galaxies
    * ....

A simple Real Life example

* Lets take an example from Downes et al. (2015)
* We have a sample with a total of N_{VLMS}=77 very low mass stars (VLMS) and N_{BD}=16 brown dwarfs (BD) from the 25 Ori cluster (~10 Myr)
* Out of these, 6 VLMS and 4 BDs have disks (infrared excesses observed)
* The key scientific question is

Do VLMS and BDs have the same disk fraction?

VLMS and BD disk Fractions

* The data are conditionally independent, thus from the product rule we have

$$P(f_{disk}^{VLMS}, f_{disk}^{BD}|data) = P(f_{disk}^{VLMS}|data)P(f_{disk}^{BD}|data)$$

i.e. the multiplication of the disk fraction posteriors for VLMS and BDs. Each of these is given by the Binomial distribution as in the coin example

$$P(f_{disk}|data) = f_{disk}^{N_{disk}}(1 - f_{disk})^{N-N_{disk}}$$

in this case we have assumed a uniform prior

$$P(f_{disk}^{VLMS}, f_{disk}^{BD}) = 1$$

* In this case our posterior is a two-dimensional function that depends upon f^{VL}_{disk} and f^{BD}_{disk}

* In this case our posterior is a two-dimensional function that depends upon f^{VL}_{disk} and f^{BD}_{disk}

$$P(f_{disk}^{VLMS}, f_{disk}^{BD} | data)$$

- In this case our posterior is a two-dimensional function that depends upon f^{VL}_{disk} and f^{BD}_{disk}

$$P(f^{VLMS}_{disk}, f^{BD}_{disk}|data)$$

* In this case our posterior is a two-dimensional function that depends upon $f^{VL}{}_{disk}$ and $f^{BD}{}_{disk}$

$$P(f^{VLMS}_{disk}, f^{BD}_{disk}|data)$$

* In this case our posterior is a two-dimensional function that depends upon f^{VL}_{disk} and f^{BD}_{disk}

$$P(f_{disk}^{VLMS}, f_{disk}^{BD}|data)$$

again, remember the posterior is the 'Holy Grail'

Marginalization

* If we want the posterior dependent upon just one of the parameters we need to marginalise over the other one

* For f^{BD}_{disk} we get

$$P(f^{BD}_{disk}|data) = \int P(f^{VLMS}_{disk}, f^{BD}_{disk}|data)df^{VLMS}_{disk}$$

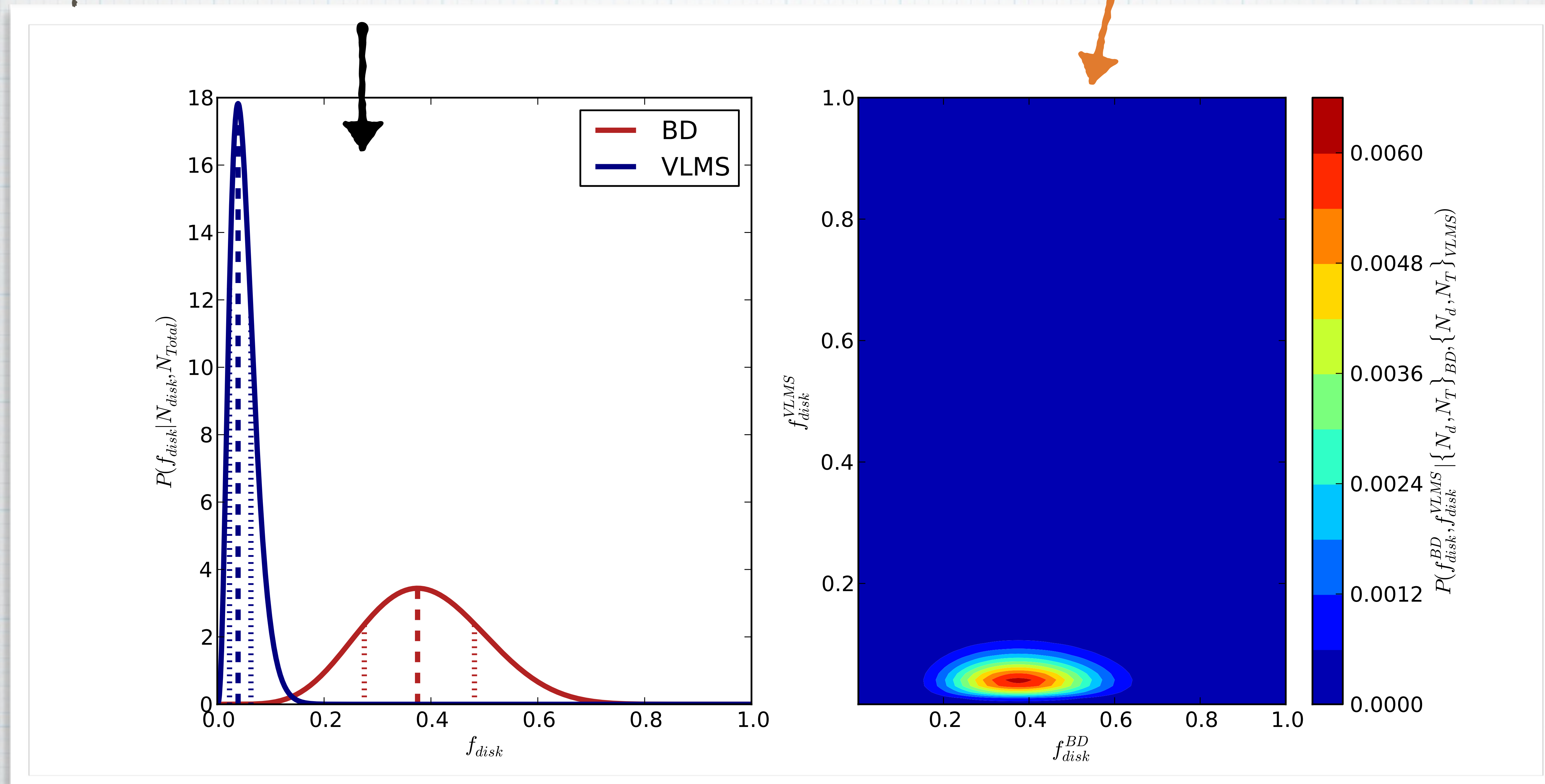
* similarly for f^{VL}_{disk}

$$P(f^{VLMS}_{disk}|data) = \int P(f^{VLMS}_{disk}, f^{BD}_{disk}|data)df^{BD}_{disk}$$

VLMS and BD disk Fractions

Marginal posteriors (integrated upon $f^{VL}{}_{disk}$ or $f^{BD}{}_{disk}$)

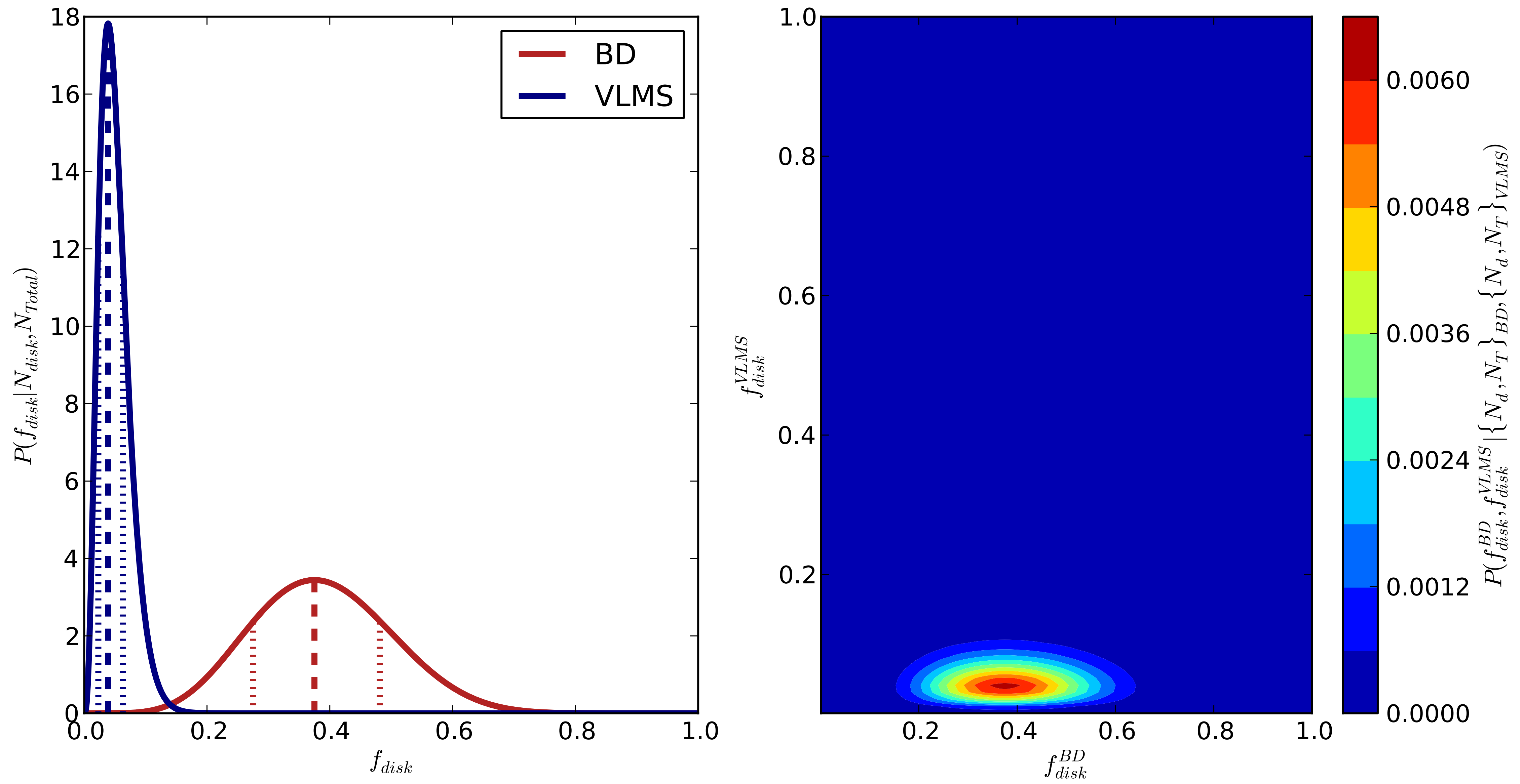
$$P(f^{VLMS}_{disk}, f^{BD}_{disk}|data)$$



$$N_{VLMS} = 77, N^{disk}_{VLMS} = 4; N^{disk}_{BD} = 16, N^{disk}_{BD} = 6$$

VLMS and BD disk Fractions

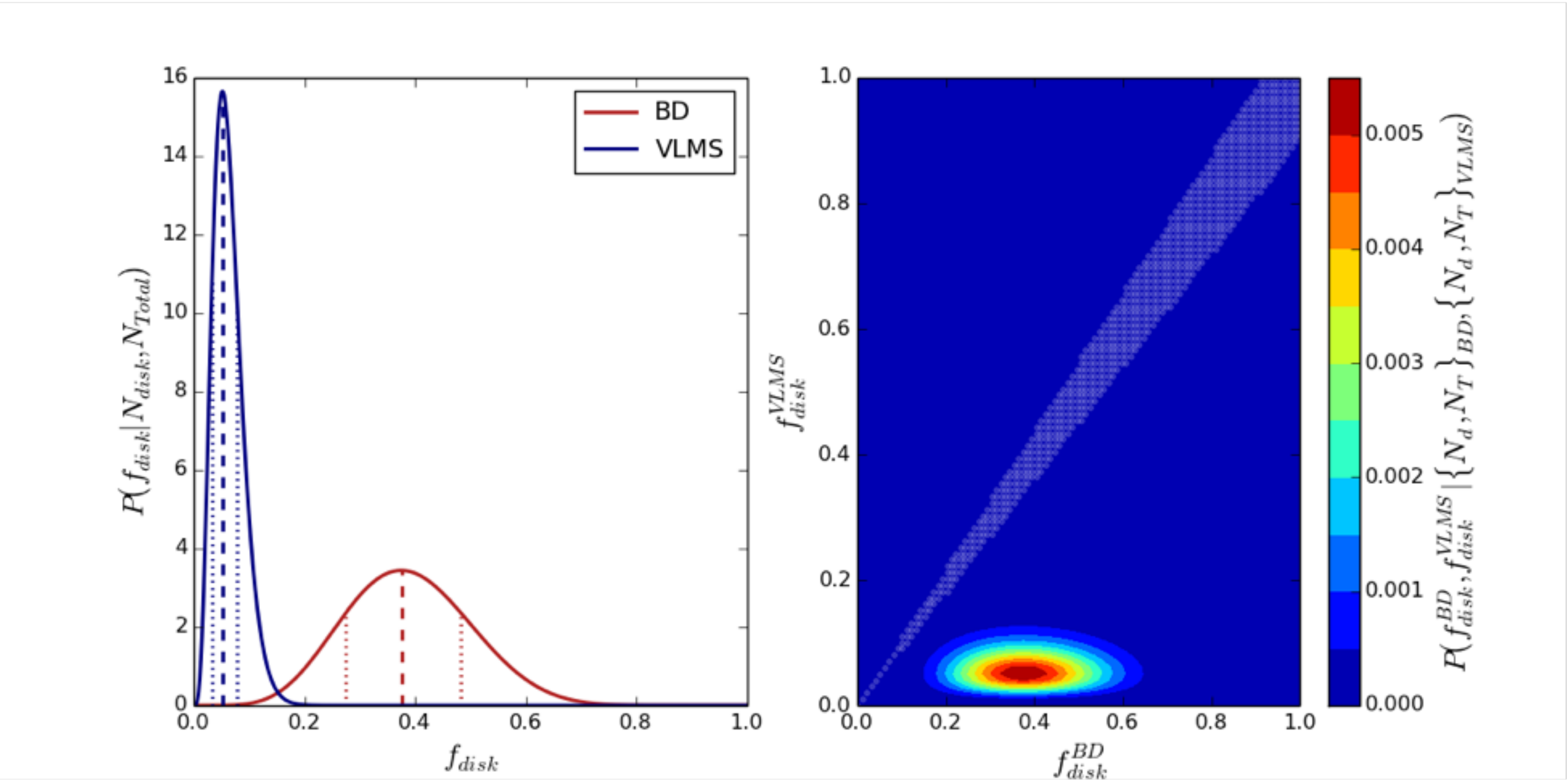
How do we answer the initial question: do VLMS and BDs have the same disk fraction?



VLMS and BD disk Fractions

Integrate the posterior in the region where

$f^{VL}_{disk} \neq f^{BD}_{disk}$

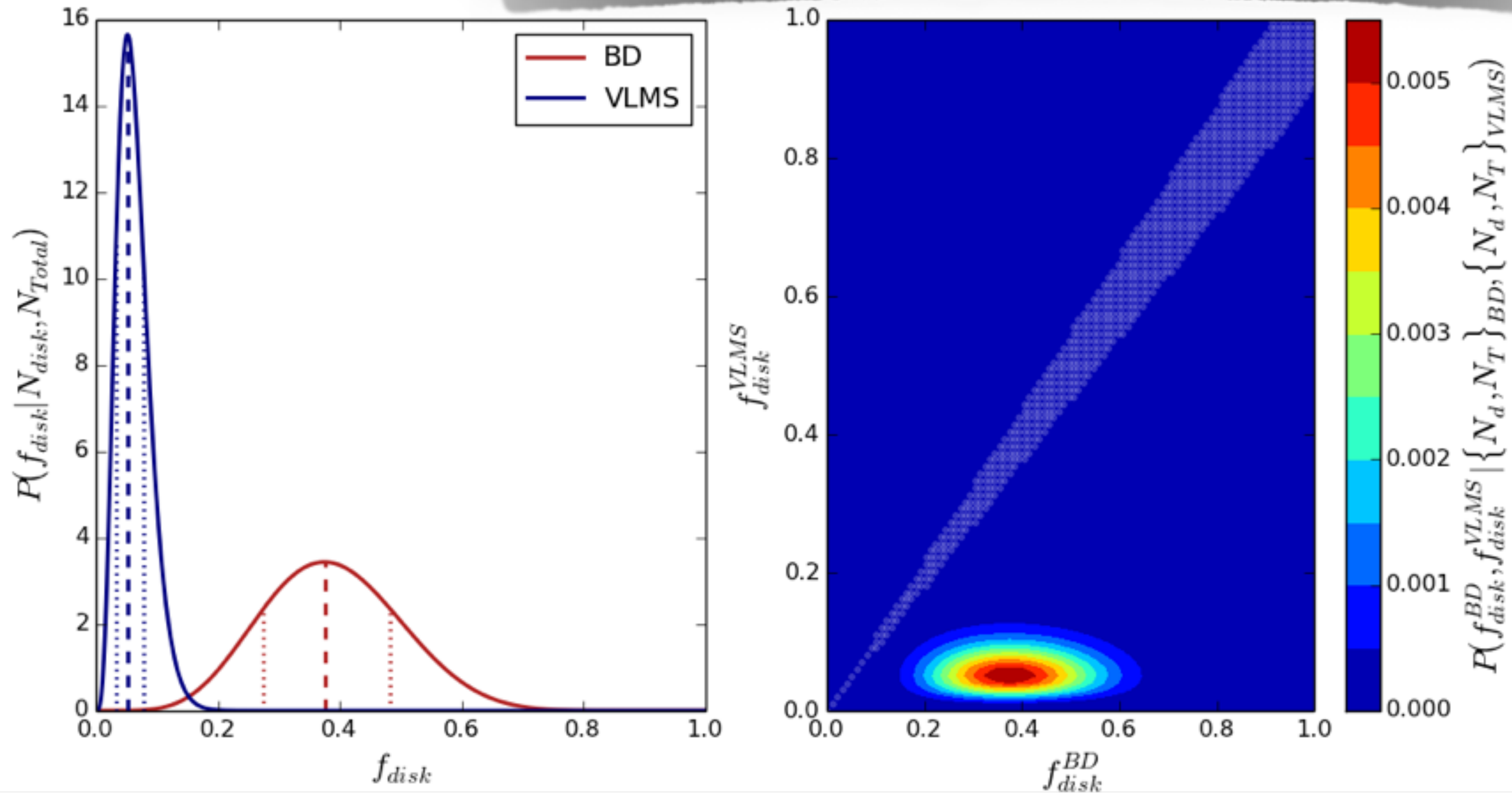


VLMS and BD disk Fractions

Integrate the posterior in the region where

$f^{VL}_{disk} \neq f^{BD}_{disk}$

this gives the probability that the two disk fractions are different!

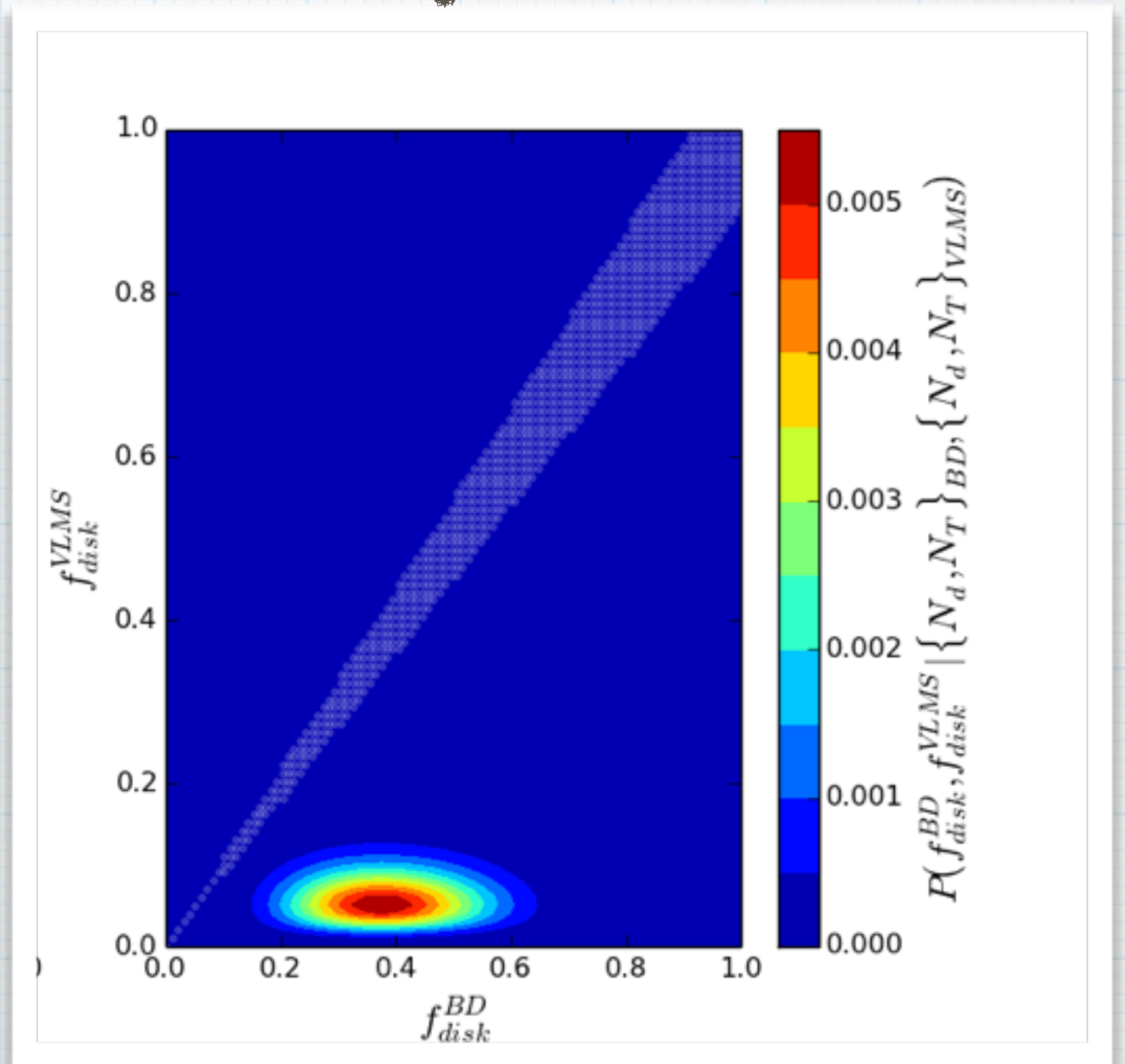


VLMS and BD disk Fractions

Better to integrate the posterior in the region where the disk fractions differ by 10% e.g.

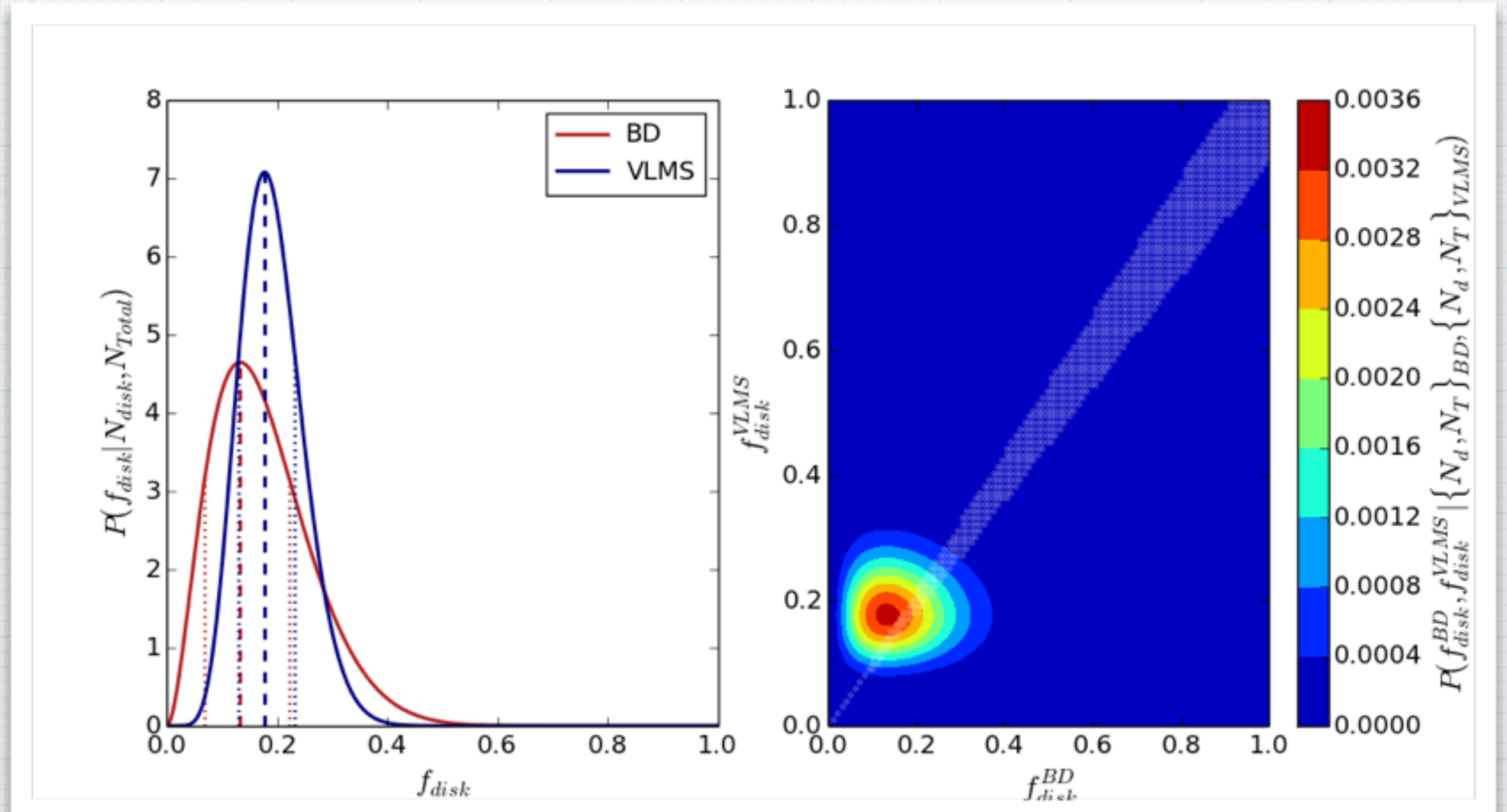
In this example we get that the probability that the two disk fractions are different is

$P_{>10\%}=99.95\%$



VLMS and BD disk Fractions

This is even more useful in more uncertain cases, e.g. for transitional disk fractions in this same cluster we have:



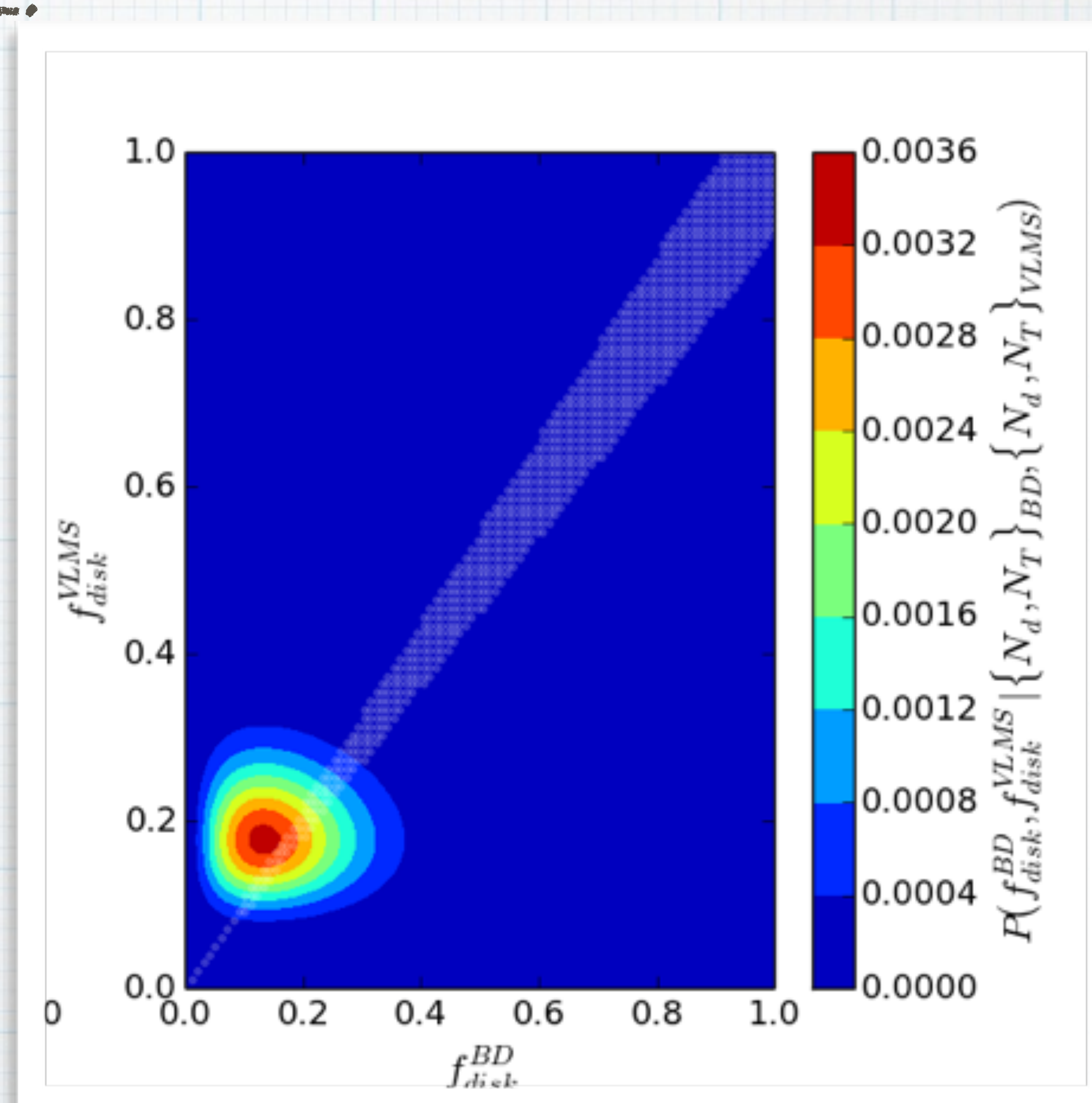
$$N_{VLMS} = 12, N^{disk}_{VLMS} = 45, N_{BD} = 8, N^{disk}_{BD} = 15$$

VLMS and BD disk Fractions

This is even more useful in more uncertain cases, e.g. for transitional disk fractions in this same cluster we have:

Here we get that the probability that the two disk fractions are different is

$P_{>10\%} = 89.7\%$



$$N_{VLMS} = 8, N^{disk}_{VLMS} = 45, N_{BD} = 2, N^{disk}_{BD} =, 15$$

Least Squares

Gaussian Uncertainties: Least Squares derived

* In a problem, whatever it may be, where our model differs from our observations because of gaussian uncertainties, the posterior is simply:

Gaussian Uncertainties: Least Squares derived

* In a problem, whatever it may be, where our model differs from our observations because of gaussian uncertainties, the posterior is simply:

$$P(model|data, I) = \prod_{i=1}^{N} e^{-\frac{(x_i - x_{model})^2}{2\sigma_i^2}} P(model|I)$$

Gaussian Uncertainties: Least Squares derived

* In a problem, whatever it may be, where our model differs from our observations because of gaussian uncertainties, the posterior is simply:

$$P(model|data, I) = \prod_{i=1}^{N} e^{-\frac{(x_i - x_{model})^2}{2\sigma_i^2}} P(model|I)$$

$$= e^{-\frac{1}{2}\sum_{i=1}^{N} \frac{(x_i - x_{model})^2}{\sigma_i^2}} P(model|I)$$

Gaussian Uncertainties: Least Squares derived

* In a problem, whatever it may be, where our model differs from our observations because of gaussian uncertainties, the posterior is simply:

$$P(model|data, I) = \prod_{i=1}^{N} e^{-\frac{(x_i - x_{model})^2}{2\sigma_i^2}} P(model|I)$$

$$= e^{-\frac{1}{2}\sum_{i=1}^{N} \frac{(x_i - x_{model})^2}{\sigma_i^2}} P(model|I)$$

* For a uniform prior we have

Gaussian Uncertainties: Least Squares derived

- In a problem, whatever it may be, where our model differs from our observations because of gaussian uncertainties, the posterior is simply:

$$P(model|data, I) = \prod_{i=1}^{N} e^{-\frac{(x_i - x_{model})^2}{2\sigma_i^2}} P(model|I)$$

$$= e^{-\frac{1}{2}\sum_{i=1}^{N} \frac{(x_i - x_{model})^2}{\sigma_i^2}} P(model|I)$$

- For a uniform prior we have

$$P(model|data, I) = e^{-\frac{1}{2}\chi^2}$$

Gaussian Uncertainties: Least Squares derived

* In a problem, whatever it may be, where our model differs from our observations because of gaussian uncertainties, the posterior is simply:

$$P(model|data, I) = \prod_{i=1}^{N} e^{-\frac{(x_i - x_{model})^2}{2\sigma_i^2}} P(model|I)$$

$$= e^{-\frac{1}{2}\sum_{i=1}^{N} \frac{(x_i - x_{model})^2}{}}$$

* For a uniform prior we have

$$P(model|data, I) = e^{-\frac{1}{2}\chi^2}$$

The least squares or $\chi 2$ minimisation method derived !!!! (assumptions are explicit!)

Very short bibliography

- Highly recommended introductory bibliography:
  - Sivia & Skilling book
  - Giulio D'Agostini's notes available at Tom Loredo's BIPS web page:

http://www.astro.cornell.edu/staff/loredo/bayes/