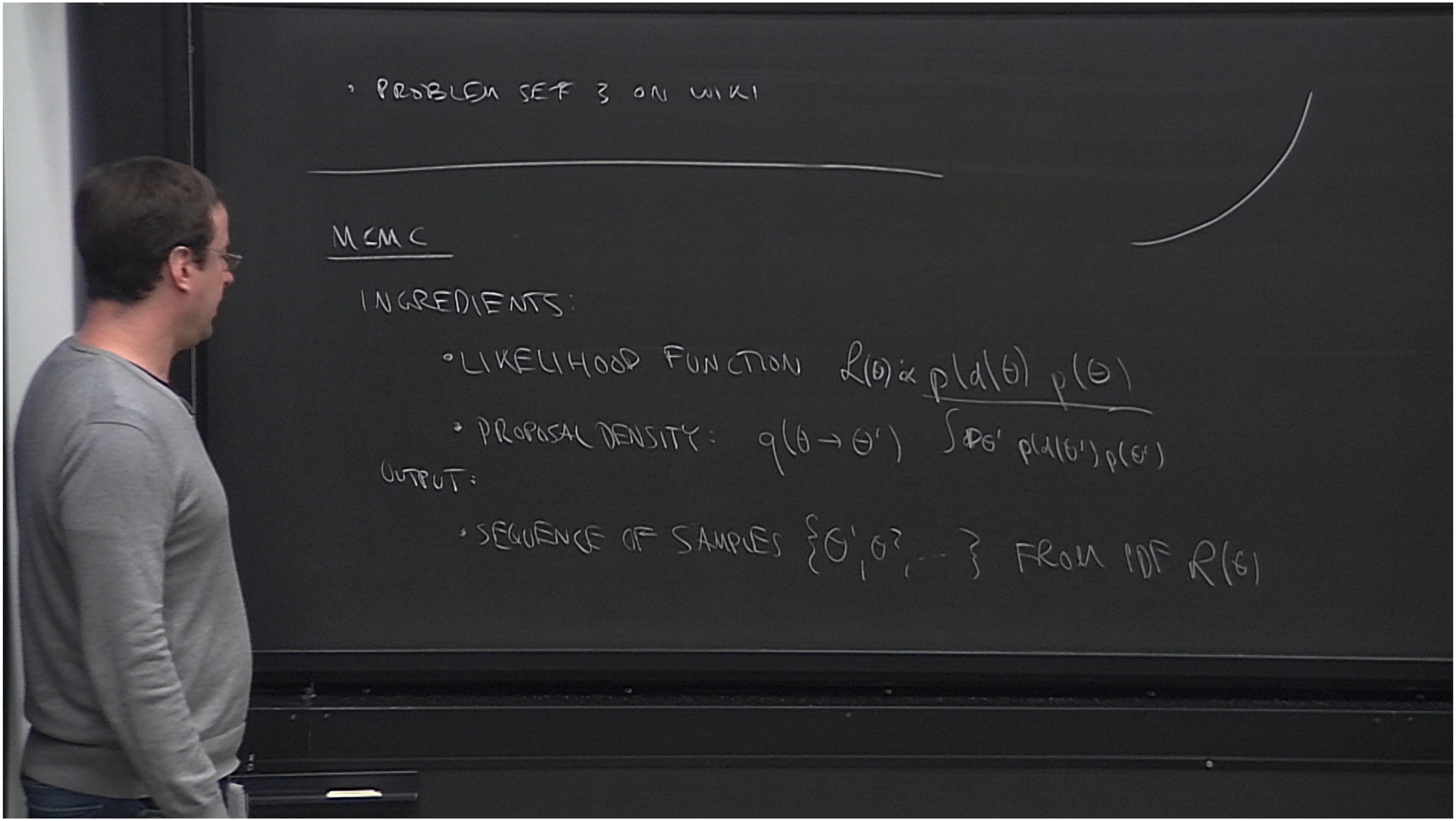
Title: PSI 2018/2019 - Explorations in Cosmology - Lecture 12

Speakers: Kendrick Smith

Collection: PSI 2018/2019 - Explorations in Cosmology (Smith)

Date: May 02, 2019 - 11:30 AM

URL: http://pirsa.org/19050006


• PROBLEM SET 3 ON WIKI
MCMC
INGREDIENTS:
• LIKELIHOOD FUNCTION $\mathcal{L}(\theta) \propto \frac{p(d|\theta)\, p(\theta)}{\int d\theta'\, p(d|\theta')\, p(\theta')}$
• PROPOSAL DENSITY: $q(\theta \to \theta')$
OUTPUT:
• SEQUENCE OF SAMPLES $\{\theta^1, \theta^2, \dots\}$ FROM PDF $\mathcal{L}(\theta)$

- PROBLEM SET 3 ON WIKI

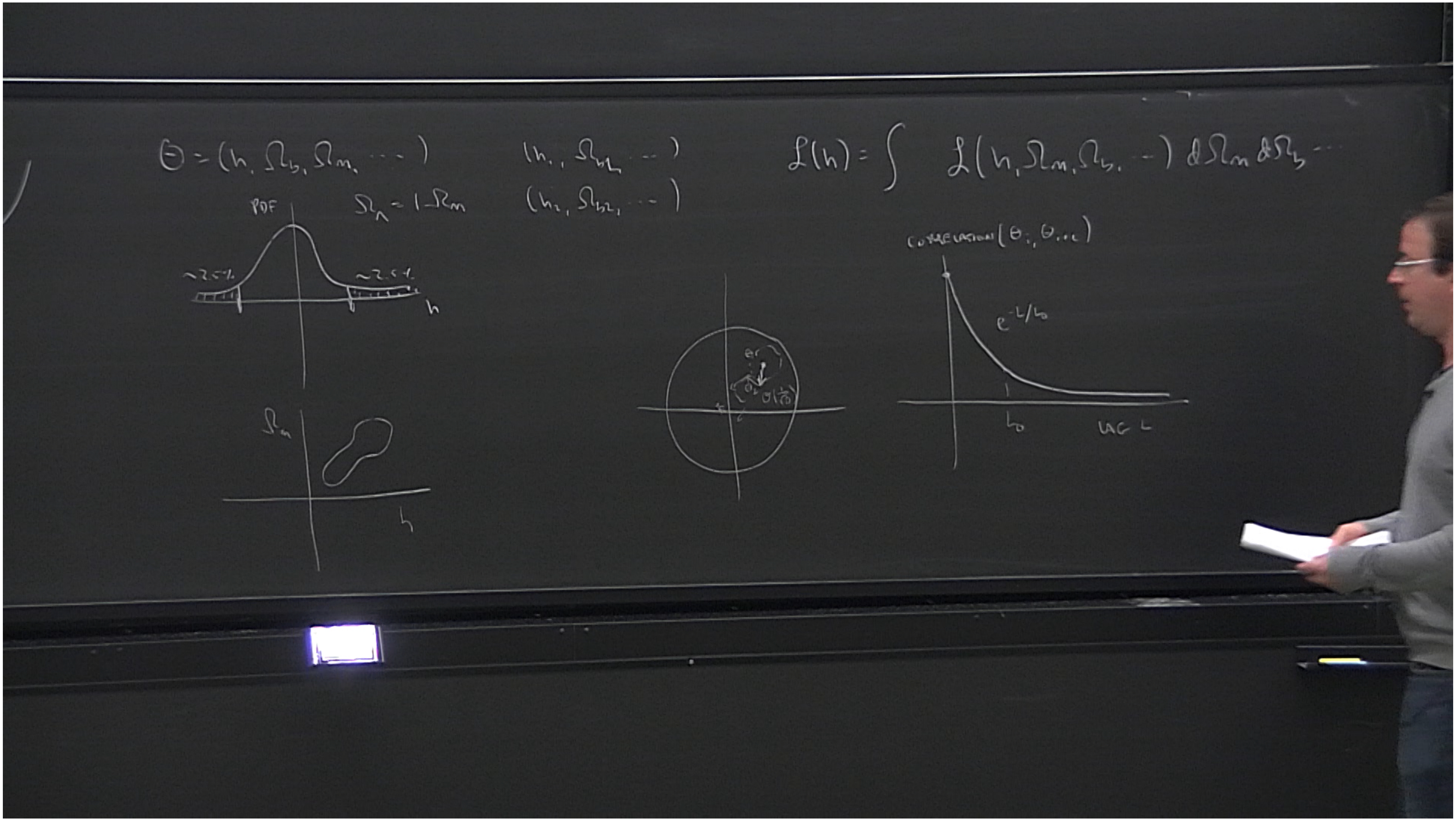
MCMC

INGREDIENTS:

- LIKELIHOOD FUNCTION $\mathcal{L}(\theta) \propto \dfrac{p(d|\theta)\, p(\theta)}{\int d\theta'\, p(d|\theta')\, p(\theta')}$
- PROPOSAL DENSITY: $q(\theta \to \theta')$

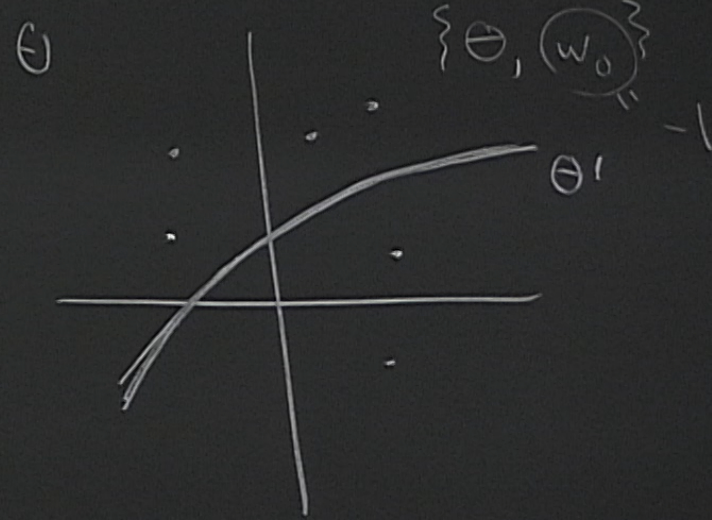
OUTPUT:

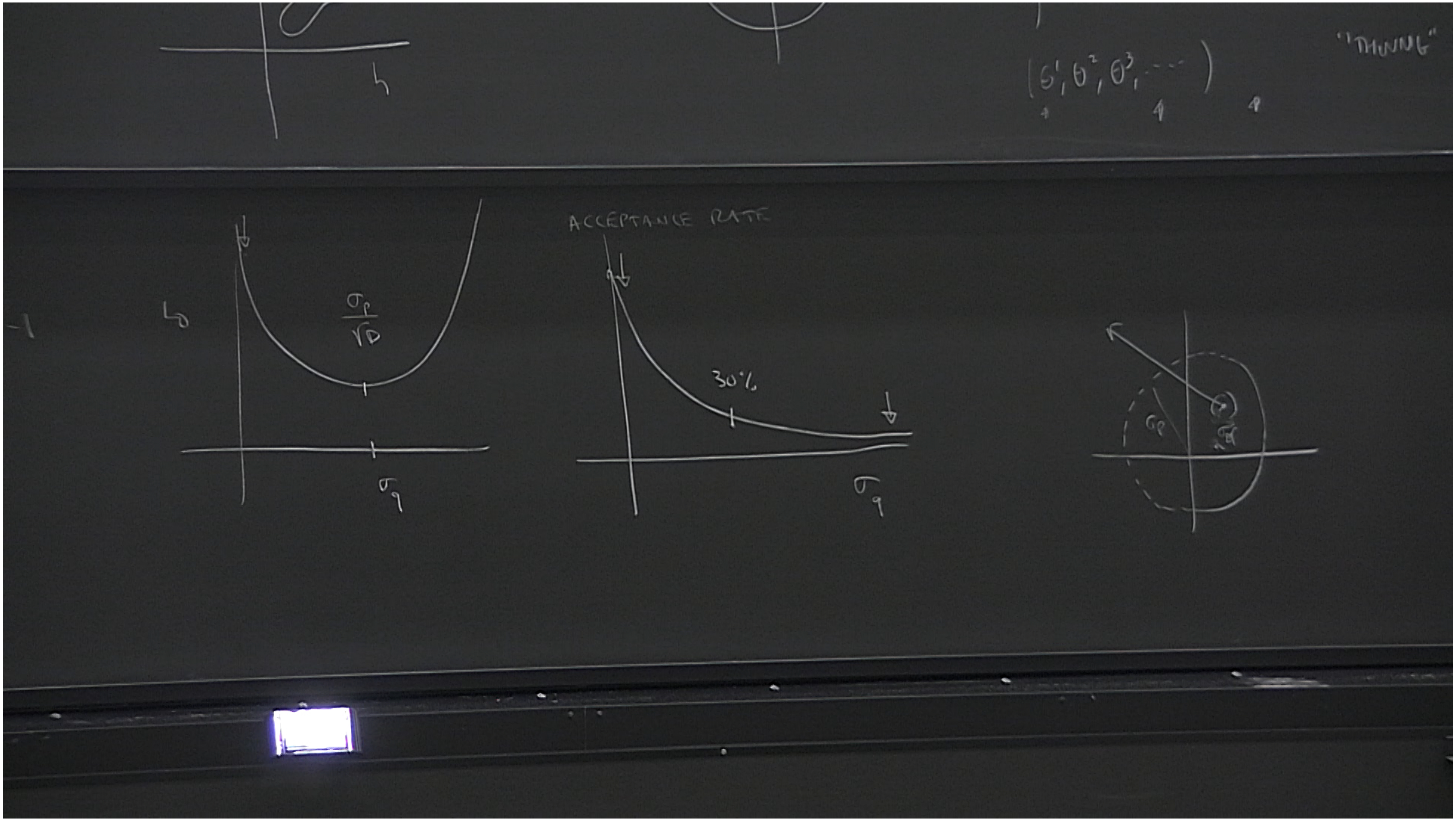
- SEQUENCE OF SAMPLES $\{\theta^1, \theta^2, \ldots\}$ FROM PDF $\mathcal{L}(\theta)$
- CAN "PROJECT" INTO 1-D AND 2-D PARAMETER SPACES

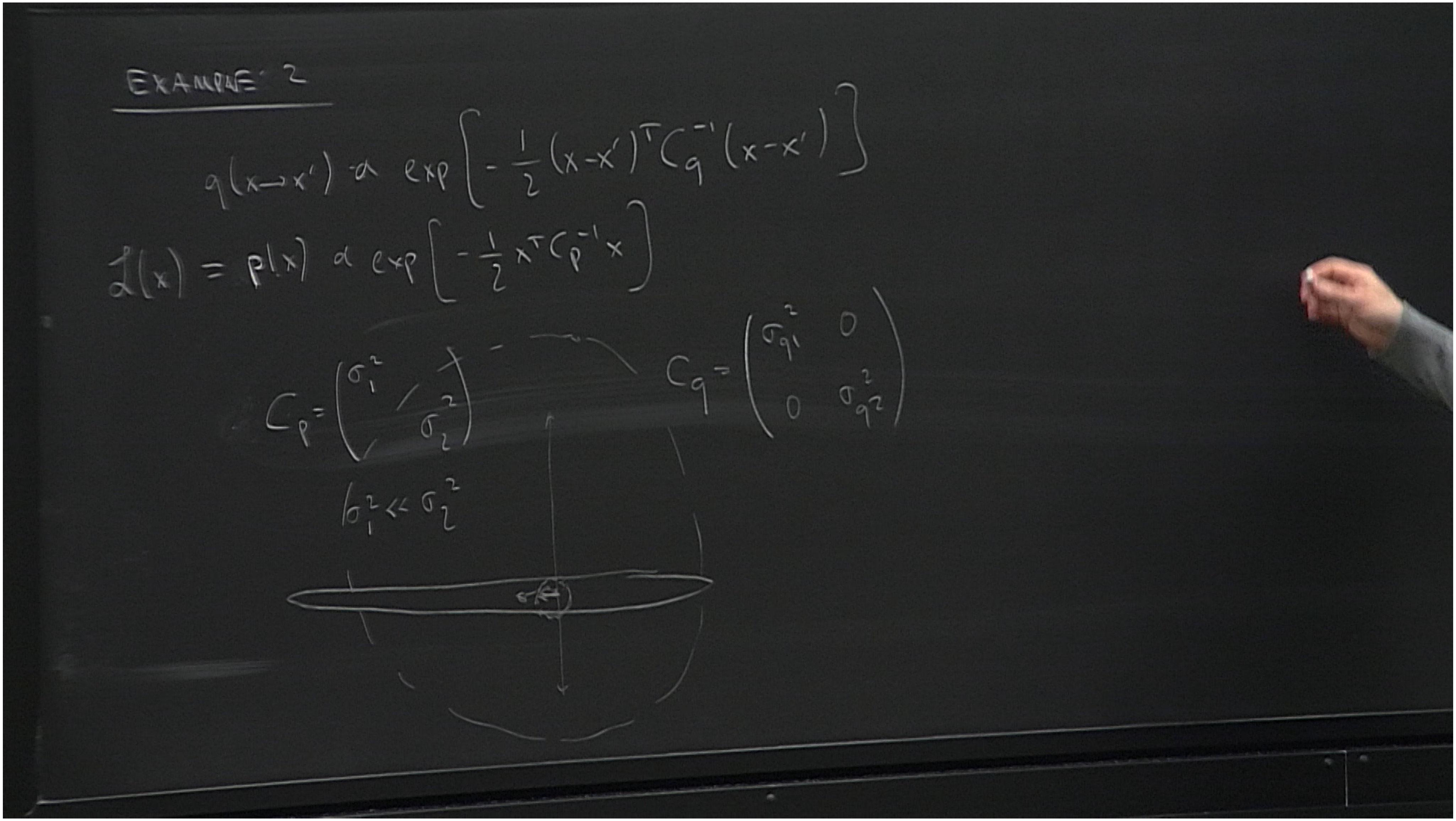

PDF
~2.5%
~2.5%
h
e^{-L/L_0}
L_0
LAG L

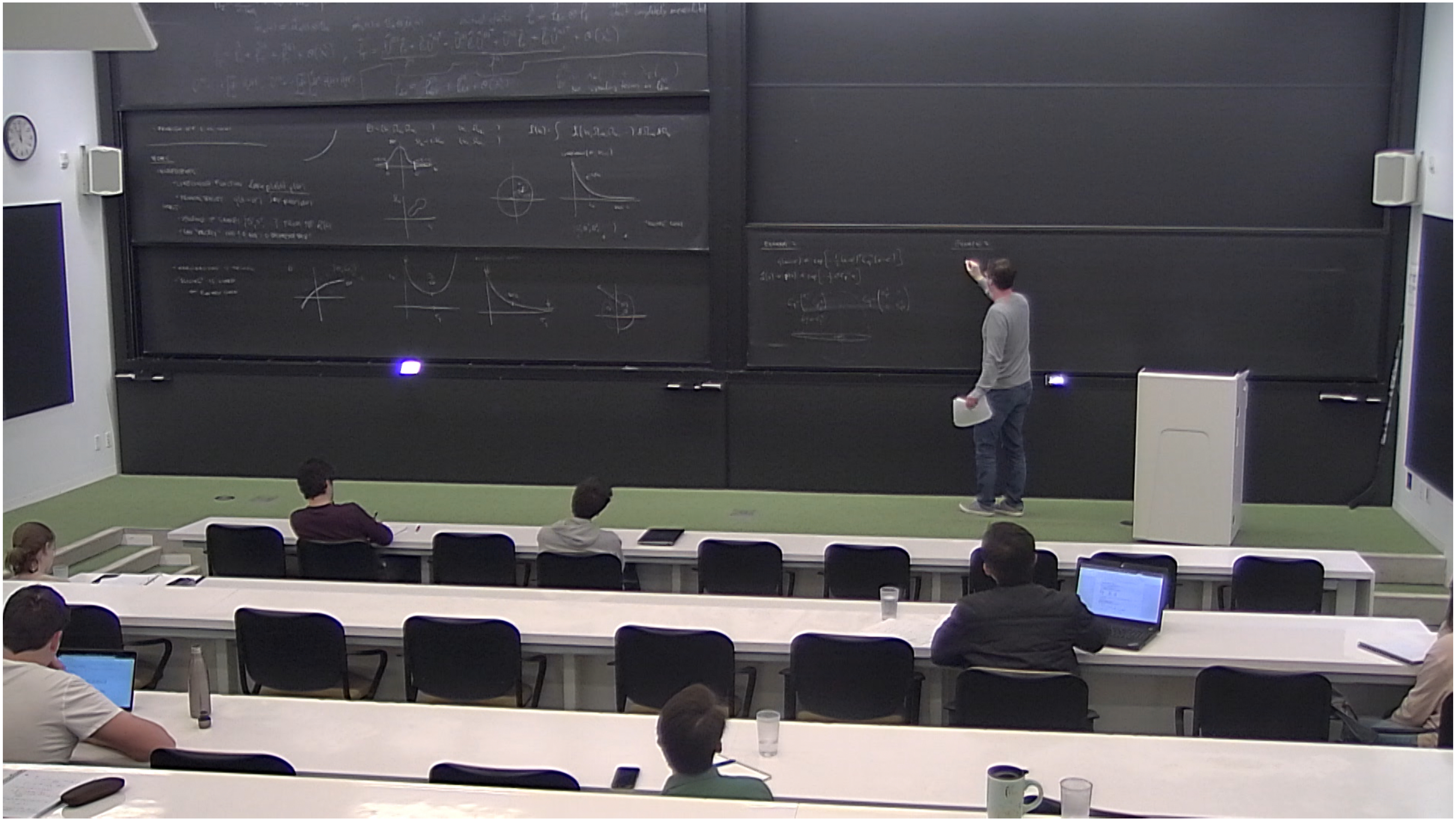
- CAN "PROJECT" INTO 1-D AND 2-D PARAMETER SPACES

- MARGINALIZING IS TRIVIAL
- "SLICING" IS HARD
  - ⟺ RUN NEW CHAIN




ACCEPTANCE RATE
30%


EXAMPLE 2
$q(x \to x') \propto \exp\left[-\frac{1}{2}(x-x')^T C_q^{-1}(x-x')\right]$
$\mathcal{L}(x) = p(x) \propto \exp\left[-\frac{1}{2}x^T C_p^{-1} x\right]$
$C_p = \begin{pmatrix} \sigma_1^2 & \\ & \sigma_2^2 \end{pmatrix}$
$C_q = \begin{pmatrix} \sigma_{q1}^2 & 0 \\ 0 & \sigma_{q2}^2 \end{pmatrix}$
$\sigma_1^2 \ll \sigma_2^2$



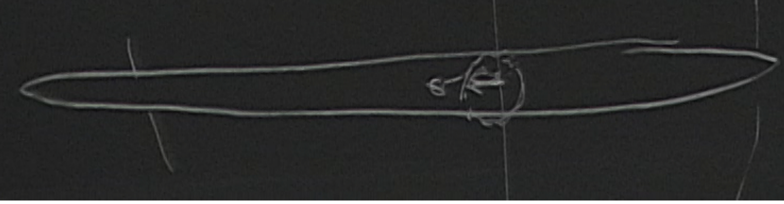
EXAMPLE 2

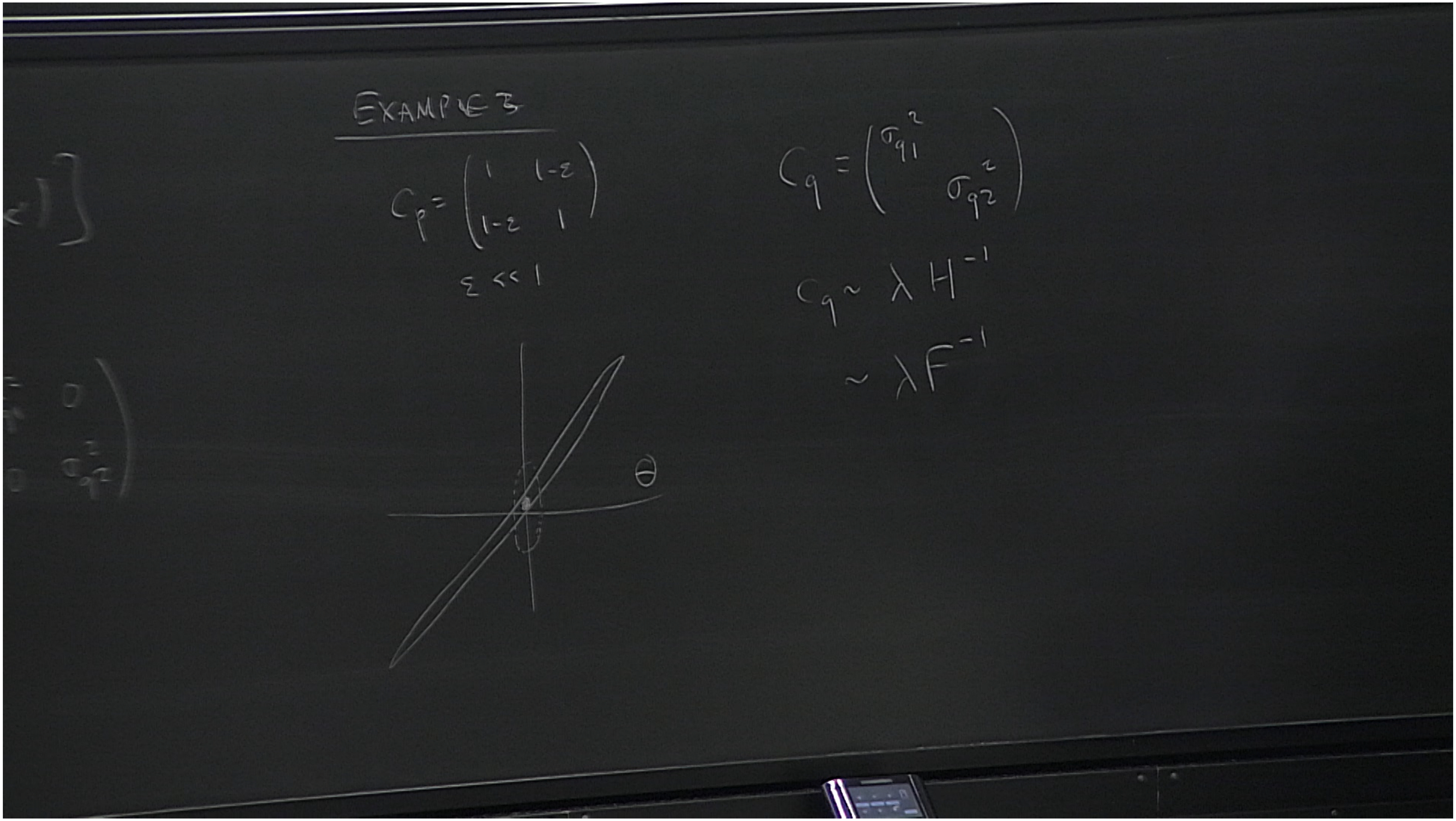
$$q(x \to x') \propto \exp\left[-\frac{1}{2}(x-x')^T C_q^{-1} (x-x')\right]$$

$$\mathcal{L}(x) = p(x) \propto \exp\left[-\frac{1}{2} x^T C_p^{-1} x\right]$$

$$C_p = \begin{pmatrix} \sigma_1^2 & \\ & \sigma_2^2 \end{pmatrix} \qquad C_q = \begin{pmatrix} \sigma_{q1}^2 & 0 \\ 0 & \sigma_{q2}^2 \end{pmatrix}$$

$$\sigma_1^2 \ll \sigma_2^2$$

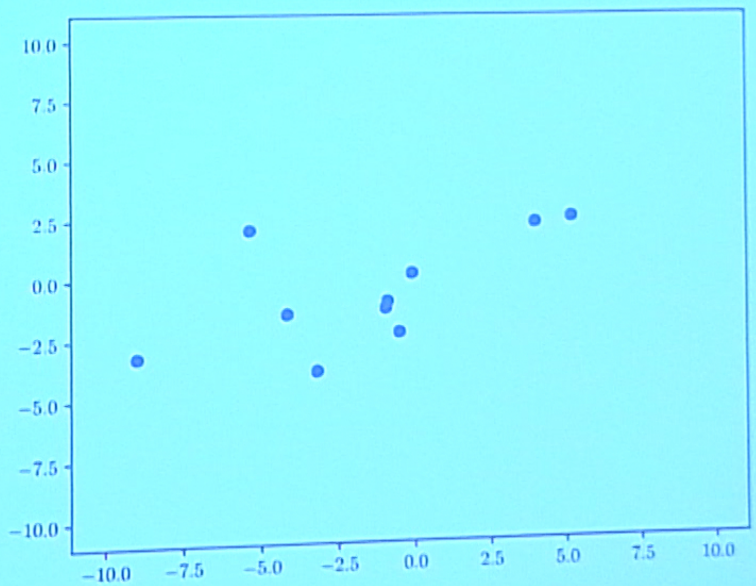



EXAMPLE 3
$C_p = \begin{pmatrix} 1 & 1-\varepsilon \\ 1-\varepsilon & 1 \end{pmatrix}$
$\varepsilon \ll 1$
$C_q = \begin{pmatrix} \sigma_{q1}^2 & \\ & \sigma_{q2}^2 \end{pmatrix}$
$C_q \sim \lambda H^{-1}$
$\sim \lambda F^{-1}$
θ

Toy MCMC example. Consider a 2-d Gaussian random variable with covariance matrix.

$$\begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}$$

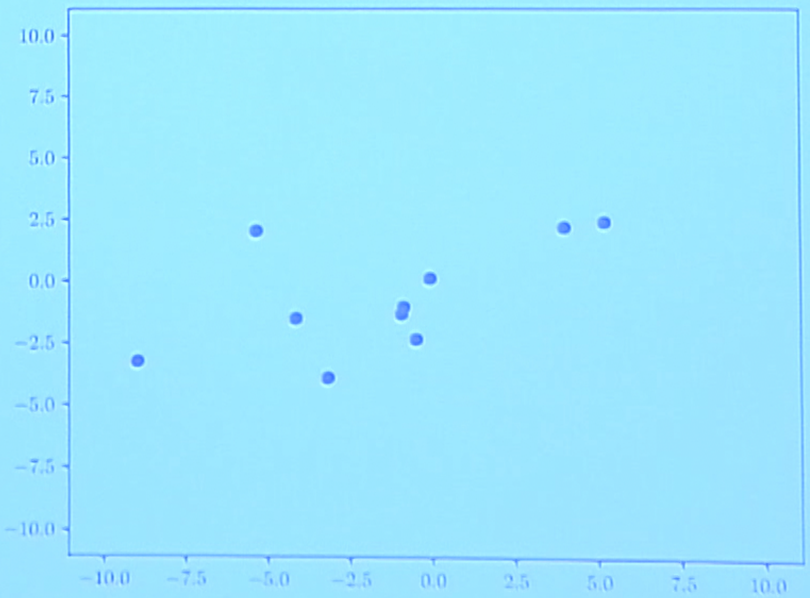
Here are 10 random samples, generated with $(\alpha,\beta) = (10,5)$.



Toy MCMC example. Consider a 2-d Gaussian random variable with covariance matrix.

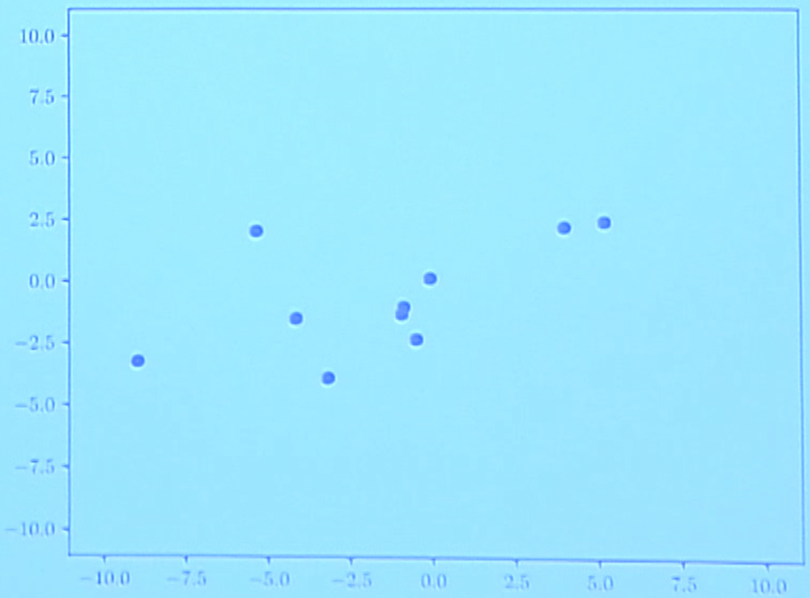
$$\begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}$$

Here are 10 random samples, generated with $(\alpha,\beta) = (10,5)$.



Now suppose you were given these 10 Gaussian samples as “data”, and wanted to know which covariance matrix produced them.

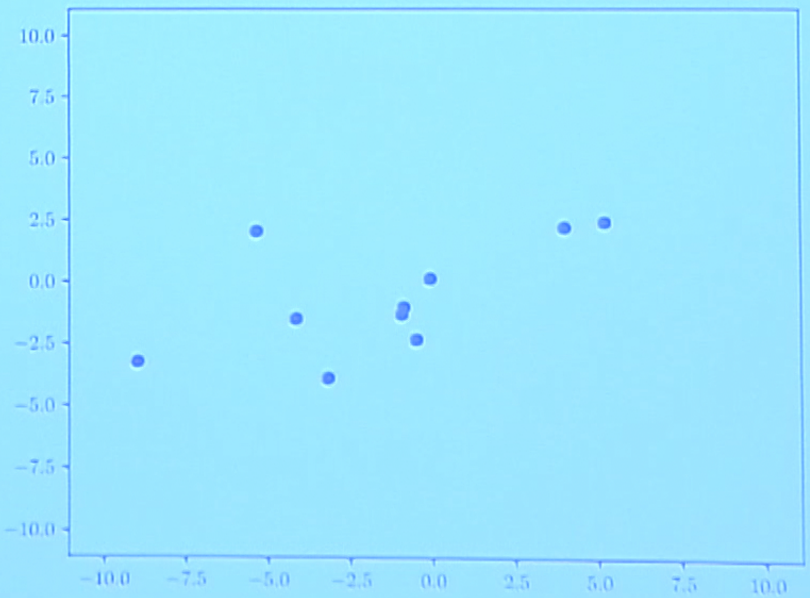
$$\begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}$$



Model parameters
$\theta^a = (\alpha, \beta)$

→

Data $d_i = (x_i, y_i)$
$i = 1, \ldots, 10$



Model parameters
$\theta^a = (\alpha, \beta)$ → Data $d_i = (x_i, y_i)$
$i = 1, \ldots, 10$

- Assume uniform prior $p(\alpha, \beta) = \text{const.}$
- Conditional likelihood $p(d|\theta)$

$$p(d|\alpha,\beta) = \prod_{i=1}^{n} \frac{1}{\mathrm{Det}(2\pi C_{\alpha,\beta})^{1/2}} \exp(-d_i^T C_{\alpha,\beta}^{-1} d_i/2)$$

where $C_{\alpha,\beta} = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}$ and $d_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix}$

Model parameters $\theta^a = (\alpha, \beta)$ → Data $d_i = (x_i, y_i)$, $i = 1, \ldots, 10$

- Assume uniform prior $p(\alpha, \beta) = \text{const.}$
- Conditional likelihood $p(d|\theta)$

$$p(d|\alpha,\beta) = \prod_{i=1}^{n} \frac{1}{\mathrm{Det}(2\pi C_{\alpha,\beta})^{1/2}} \exp(-d_i^T C_{\alpha,\beta}^{-1} d_i/2)$$

where $C_{\alpha,\beta} = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}$ and $d_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix}$

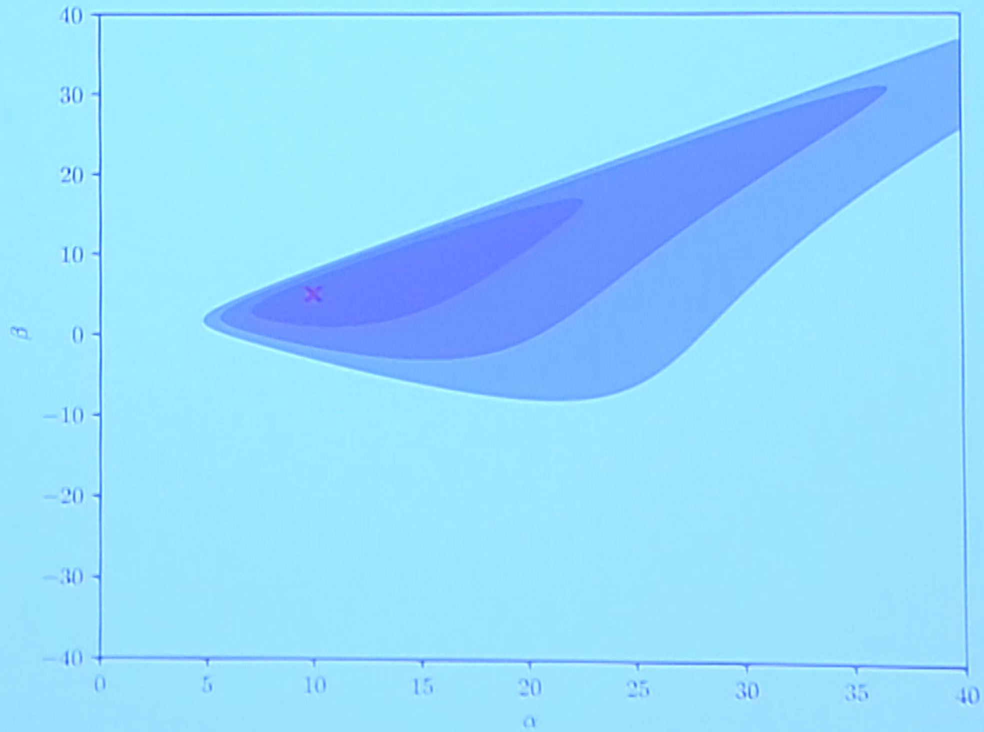
- Posterior likelihood $p(\theta|d) \propto p(d|\theta)p(\theta)$

$$p(\alpha,\beta|d) \propto \prod_{i=1}^{n} \frac{1}{\mathrm{Det}(2\pi C_{\alpha,\beta})^{1/2}} \exp(-d_i^T C_{\alpha,\beta}^{-1} d_i/2)$$

In this toy example, it is not really necessary to use an MCMC.

The posterior p(α,β|d) is fast to compute (10 terms) and there are only two model parameters (α,β).

Can just compute p(α,β|d) on a grid of (α,β) values and plot it!



As an aside, let's compare the Bayesian posterior to the Fisher approximation, in this toy problem.

First we need to calculate the Fisher matrix, starting from the conditional likelihood $p(d|\theta^a)$:

$$p(d|\alpha,\beta) = \prod_{i=1}^{n} \frac{1}{\mathrm{Det}(2\pi C_{\alpha,\beta})^{1/2}} \exp(-d_i^T C_{\alpha,\beta}^{-1} d_i/2)$$

This calculation gives (details omitted):

$$F_{ab} = -\left\langle \frac{\partial^2 \log(d|\theta)}{\partial\theta^a\,\partial\theta^b} \right\rangle_d = \frac{n}{2}\mathrm{Tr}\left(C_{\mathrm{fid}}^{-1} \frac{\partial C}{\partial\theta^a} C_{\mathrm{fid}}^{-1} \frac{\partial C}{\partial\theta^b} \right)$$

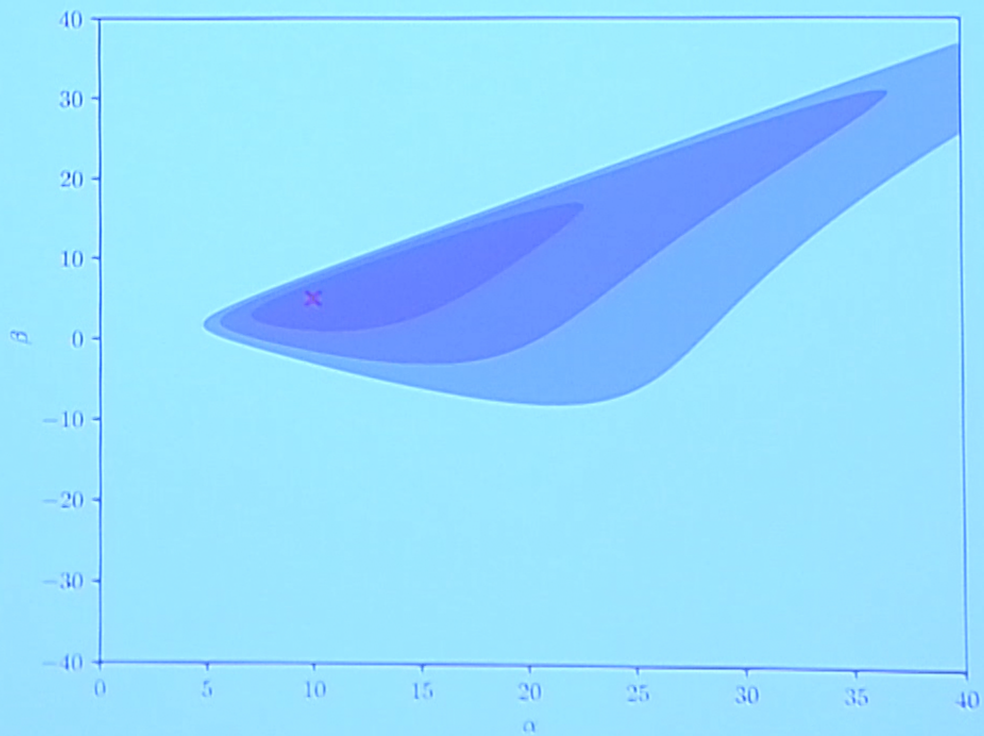
where

$$C_{\mathrm{fid}} = \begin{pmatrix} \alpha_{\mathrm{fid}} & \beta_{\mathrm{fid}} \\ \beta_{\mathrm{fid}} & \alpha_{\mathrm{fid}} \end{pmatrix} \qquad \frac{\partial C}{\partial\alpha} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \frac{\partial C}{\partial\beta} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

In this toy example, it is not really necessary to use an MCMC.

The posterior $p(\alpha,\beta|d)$ is fast to compute (10 terms) and there are only two model parameters (α,β).

Can just compute $p(\alpha,\beta|d)$ on a grid of (α,β) values and plot it!



As an aside, let's compare the Bayesian posterior to the Fisher approximation, in this toy problem.

First we need to calculate the Fisher matrix, starting from the conditional likelihood p(d|θ^a):

$$p(d|\alpha,\beta) = \prod_{i=1}^{n} \frac{1}{\mathrm{Det}(2\pi C_{\alpha,\beta})^{1/2}} \exp(-d_i^T C_{\alpha,\beta}^{-1} d_i/2)$$

This calculation gives (details omitted):

$$F_{ab} = -\left\langle \frac{\partial^2 \log(d|\theta)}{\partial\theta^a\,\partial\theta^b} \right\rangle_d$$

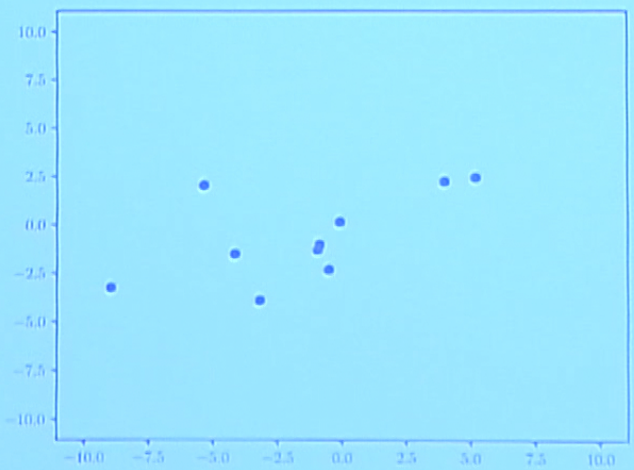
$$= \frac{n}{2}\mathrm{Tr}\left(C_{\mathrm{fid}}^{-1} \frac{\partial C}{\partial\theta^a} C_{\mathrm{fid}}^{-1} \frac{\partial C}{\partial\theta^b} \right)$$

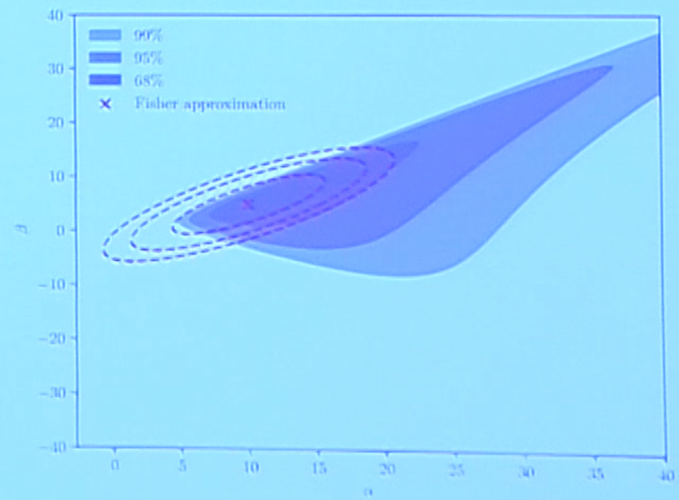
where

$$C_{\mathrm{fid}} = \begin{pmatrix} \alpha_{\mathrm{fid}} & \beta_{\mathrm{fid}} \\ \beta_{\mathrm{fid}} & \alpha_{\mathrm{fid}} \end{pmatrix} \quad \frac{\partial C}{\partial\alpha} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \frac{\partial C}{\partial\beta} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

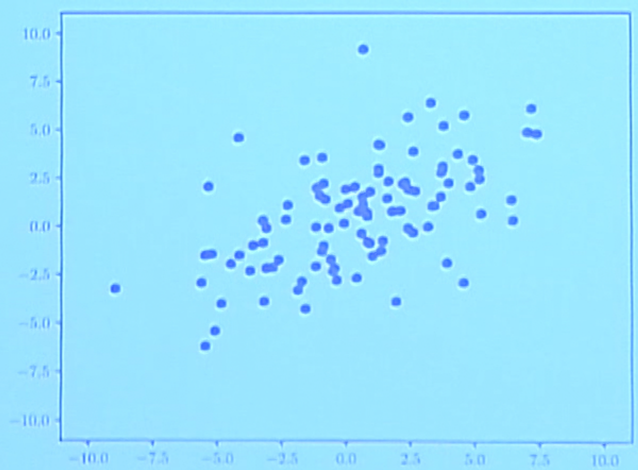
With N=10 points in the original dataset, the Fisher approximation is pretty rough.

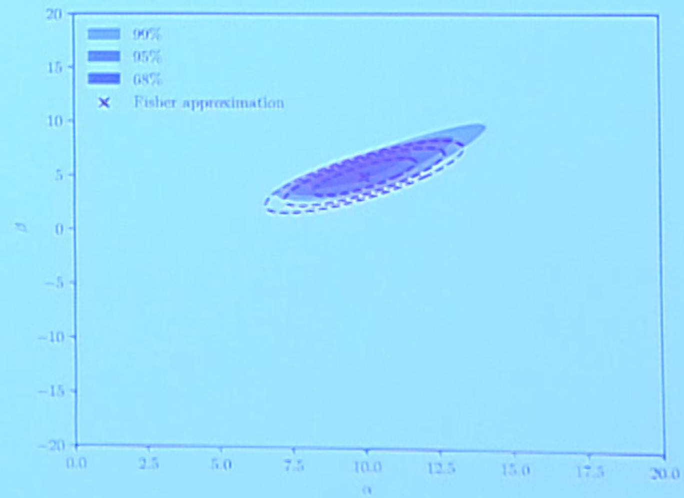
The Bayesian central limit theorem implies that it should become more accurate if N is increased.





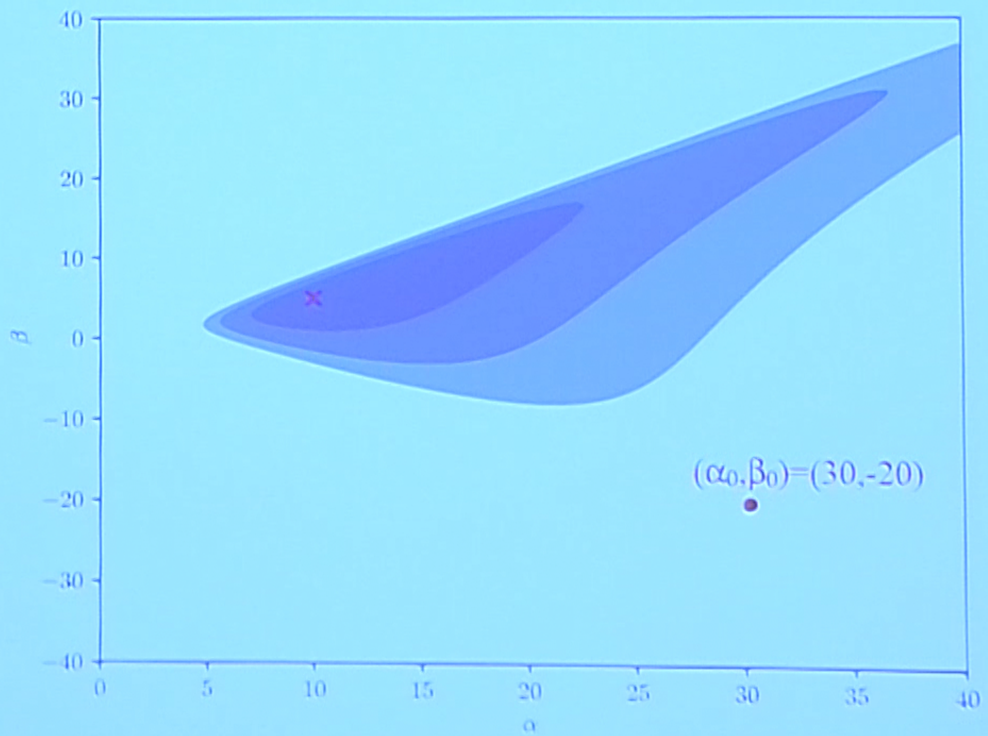
For N=100, the Fisher approximation is pretty good.





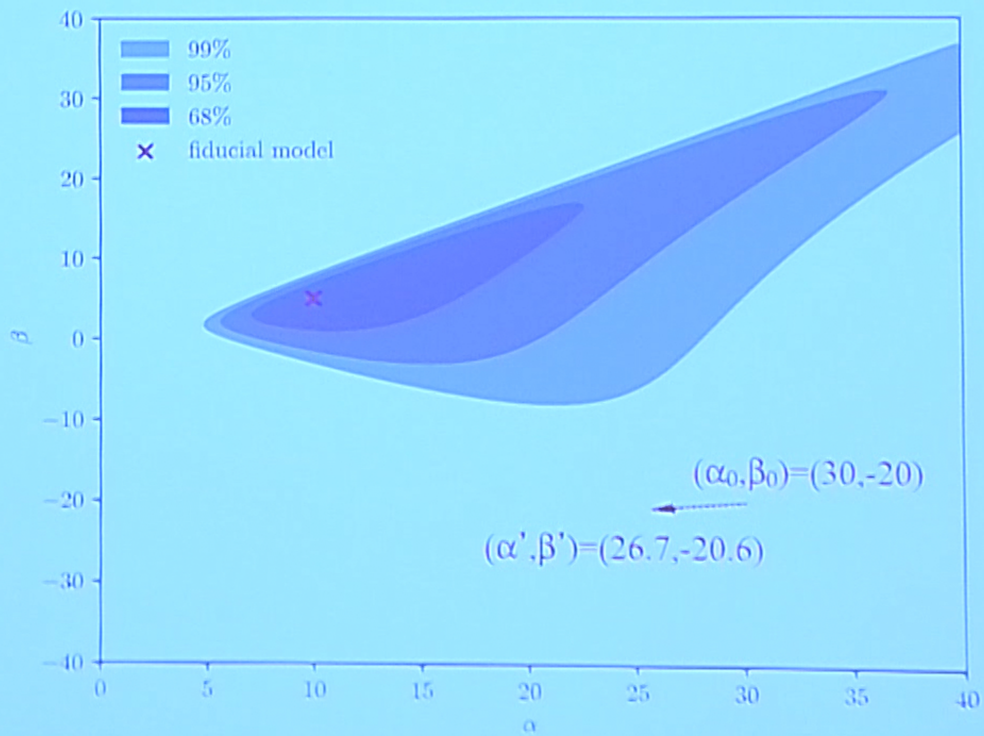
Going back to the N=10 case, let's try to reproduce the likelihood with an MCMC.

I chose $(\alpha_0,\beta_0)=(30,-20)$ for the starting location.

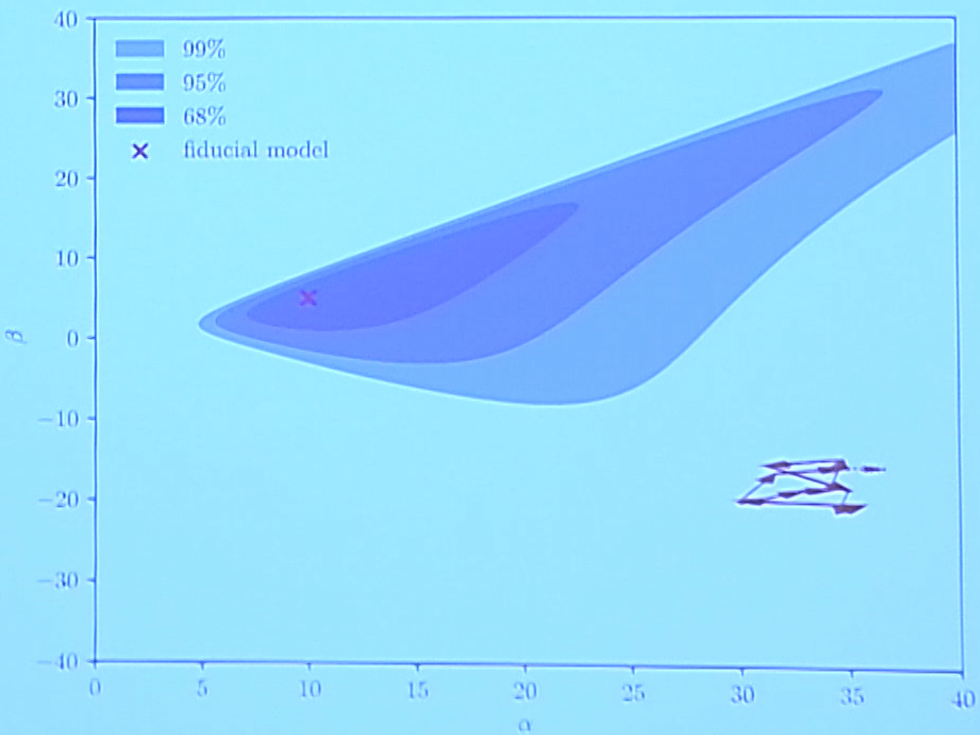


Step 1: likelihood ratio $\frac{P(\alpha', \beta' | d)}{P(\alpha_0, \beta_0 | d)} = 0.047$, reject

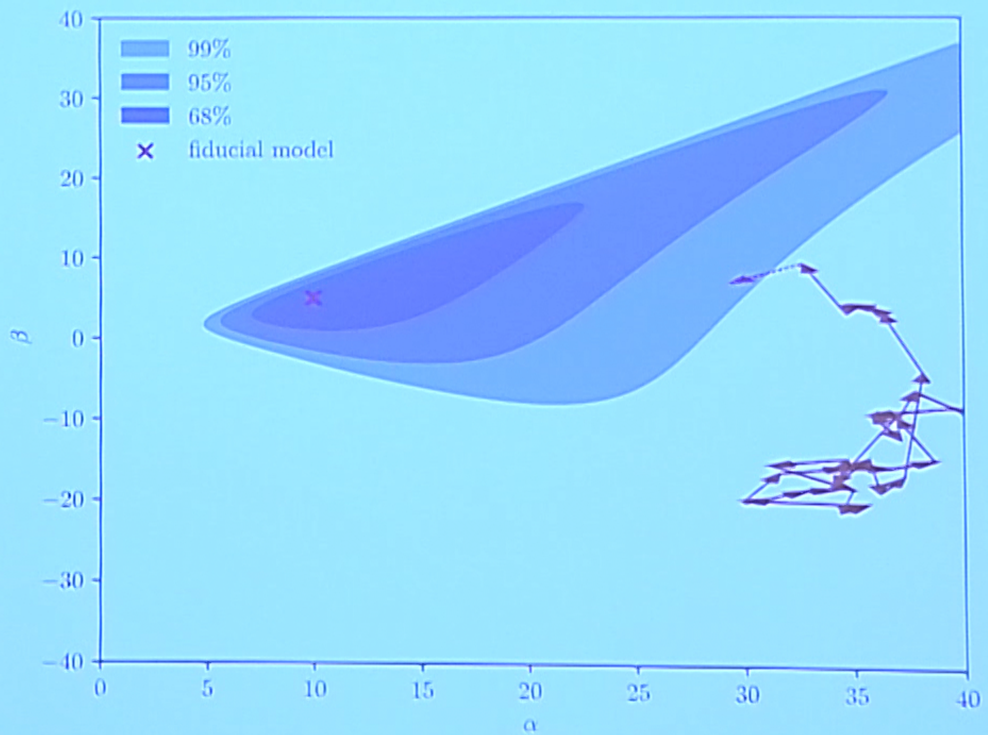
Chain so far: (α_0, β_0), (α_0, β_0)



After 20 steps (55% acceptance fraction)



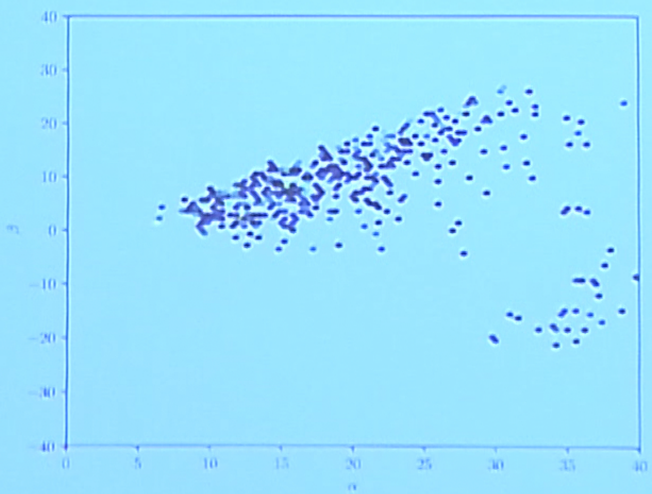
After 50 steps

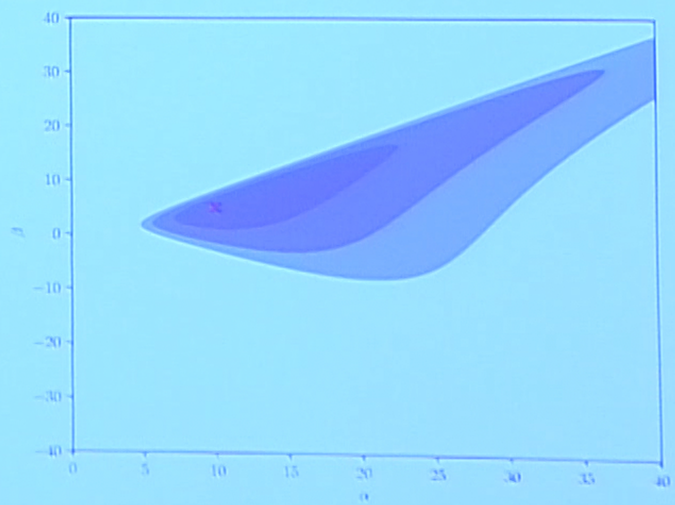


After 1000 steps (color-coding includes multiplicities)

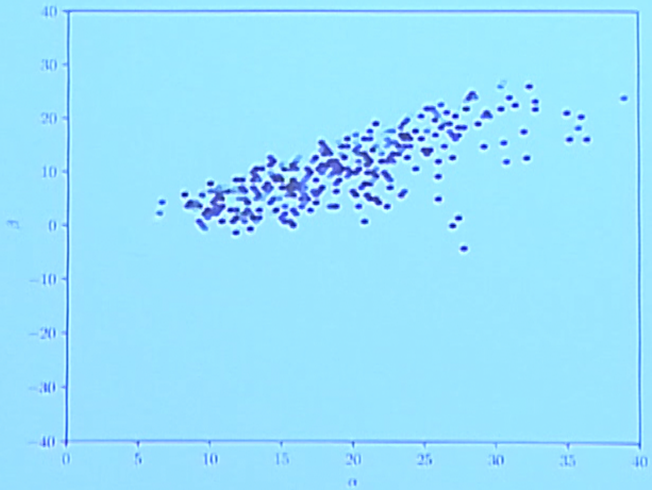
Detail: the cluster of points on the lower right is an artifact of the initial location $(\alpha_0,\beta_0)=(30,-20)$. It would be better to discard (say) the first 200 samples (the "burn-in" period).

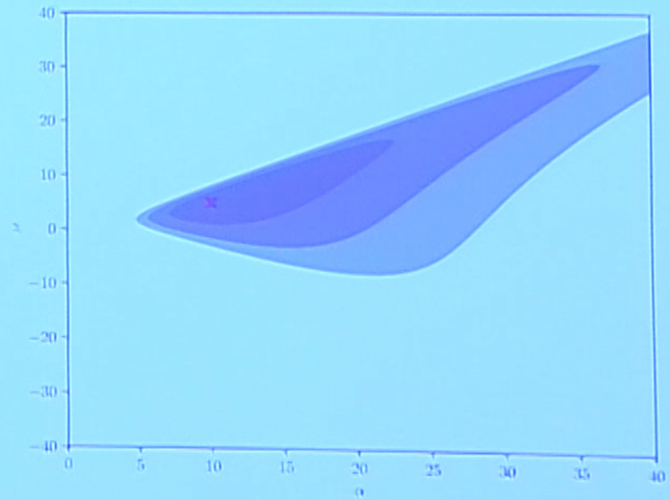
A common rule of thumb: discard the first 10% of the chain.



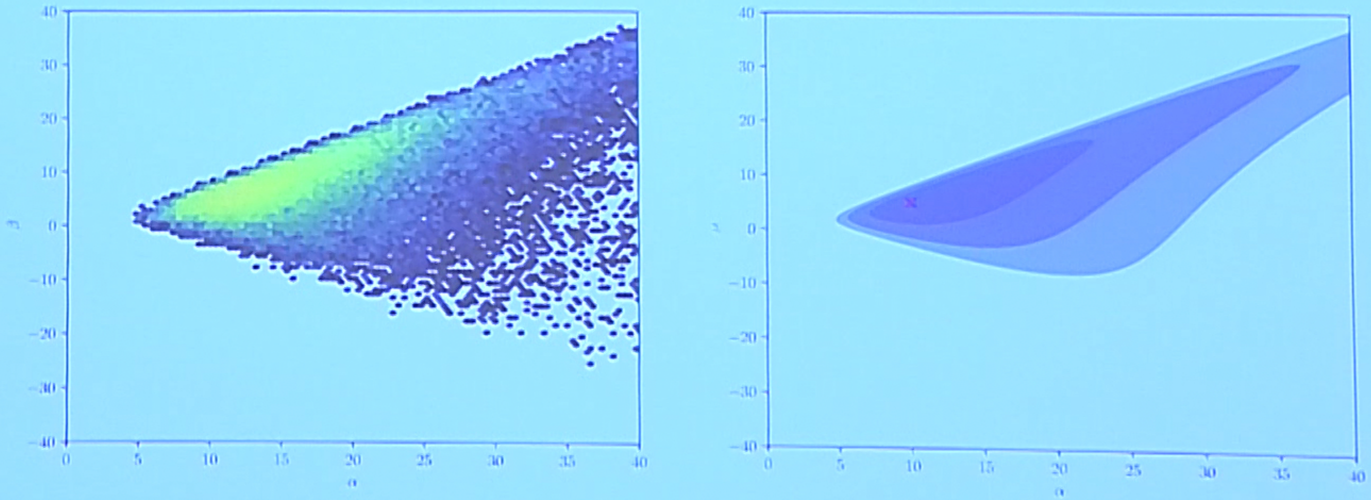


After 1000 steps, burn-in discarded.

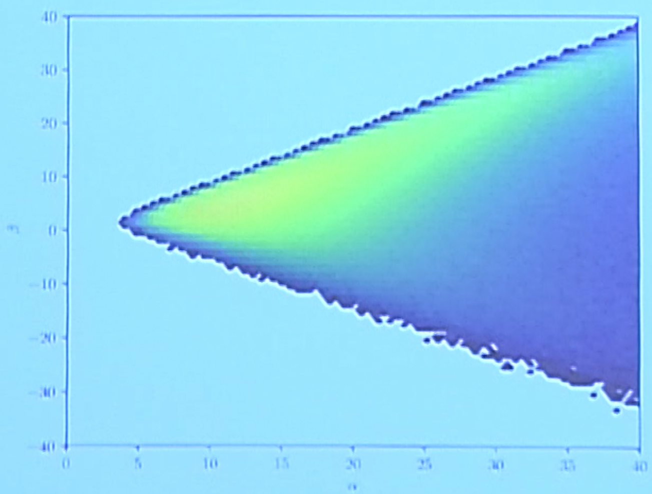


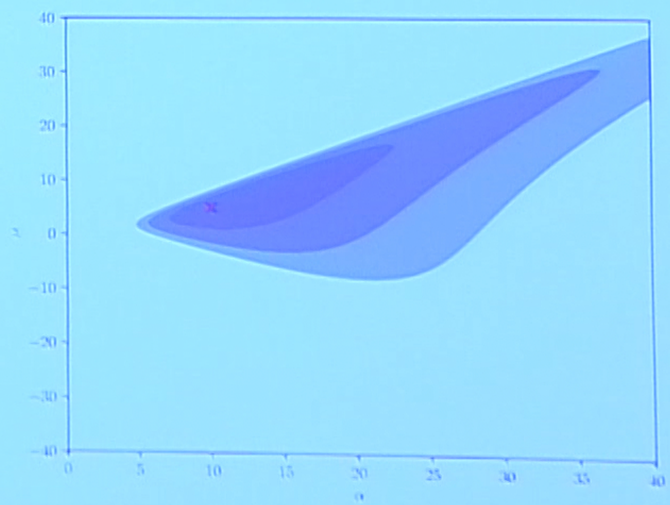


After 10^5 steps, burn-in discarded.



After 10^7 steps, burn-in discarded.

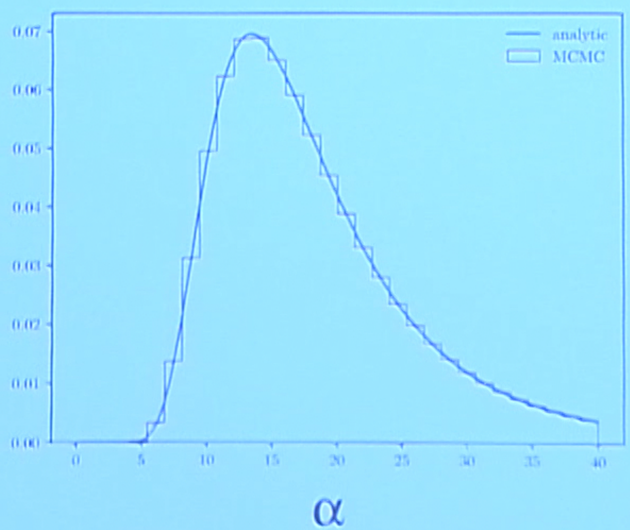


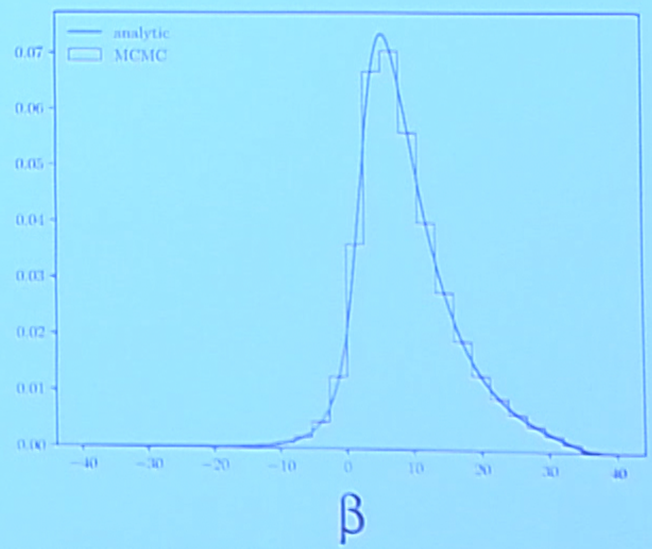


Comparing 1-d marginalized likelihoods.

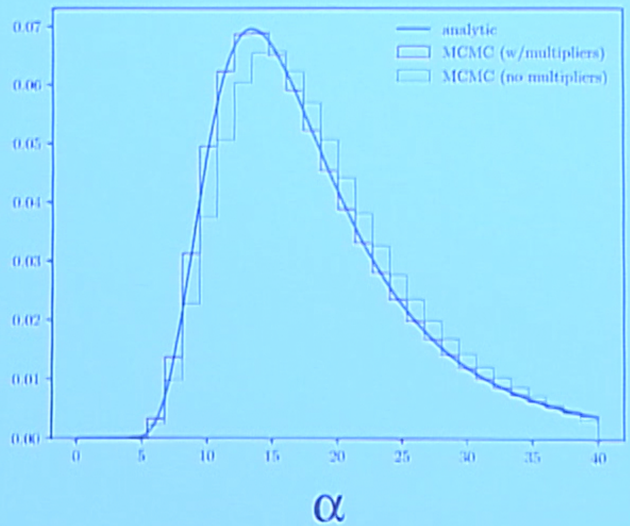
$$\mathcal{L}(\alpha) = \int d\beta\, \mathcal{L}(\alpha, \beta)$$

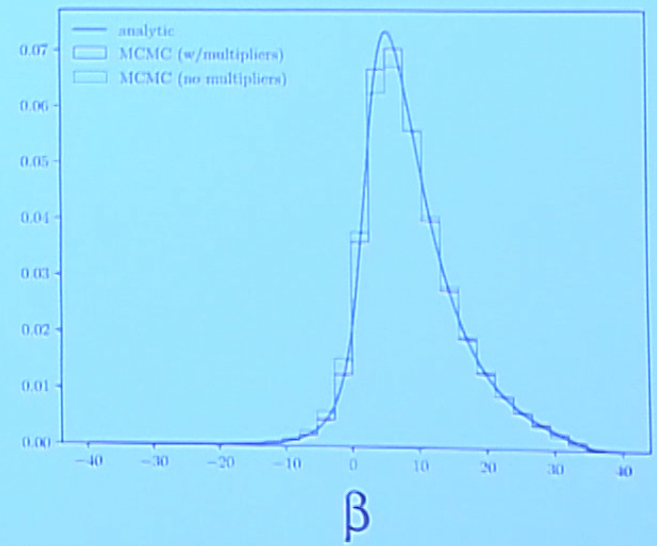
$$\mathcal{L}(\beta) = \int d\alpha\, \mathcal{L}(\alpha, \beta)$$

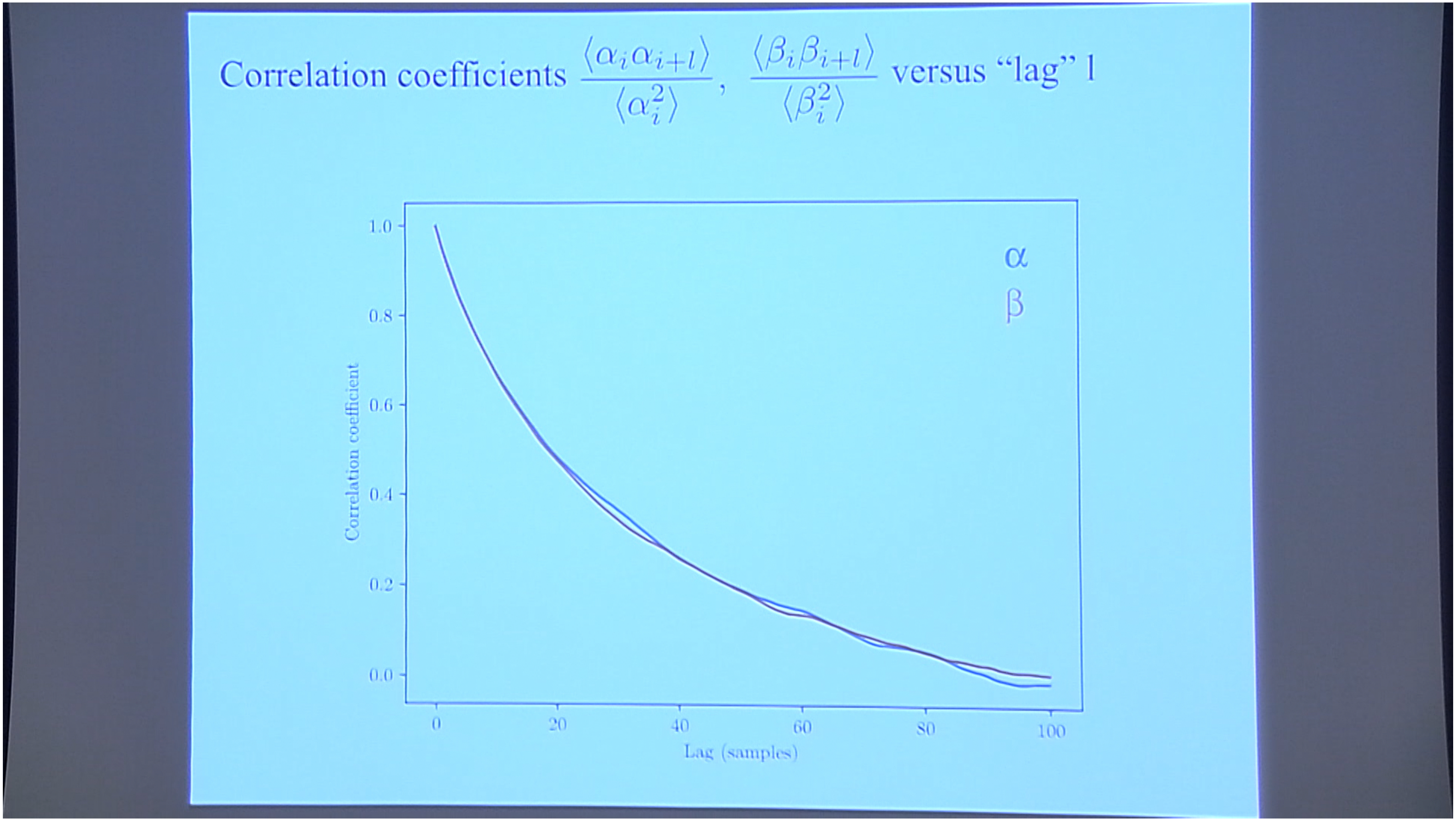




Reminder: must count MCMC samples with "multiplicity", which arises from rejection.

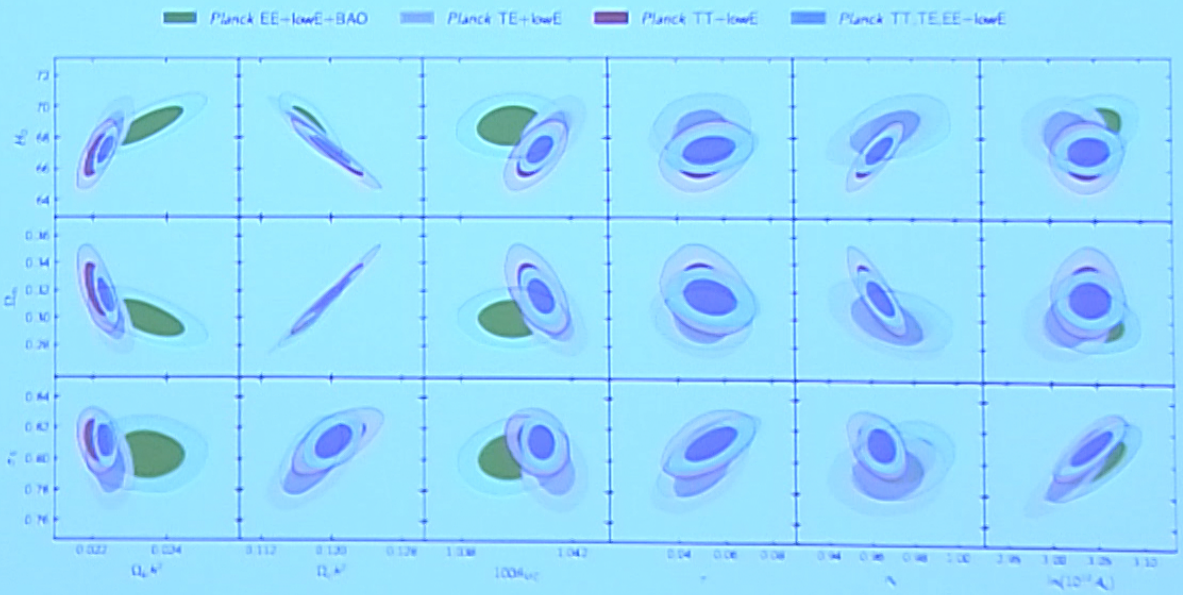





Correlation coefficients $\frac{\langle \alpha_i \alpha_{i+l} \rangle}{\langle \alpha_i^2 \rangle}$, $\frac{\langle \beta_i \beta_{i+l} \rangle}{\langle \beta_i^2 \rangle}$ versus "lag" l
α
β
Correlation coefficient
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
Lag (samples)

Cosmological parameters
+ nuisance parameters → CMB multipoles a_{lm}

Morally similar, but likelihood function $P(a_{lm} \mid \theta^a)$ is much more complicated (e.g. involves running CAMB to compute C_l)



Planck 2018