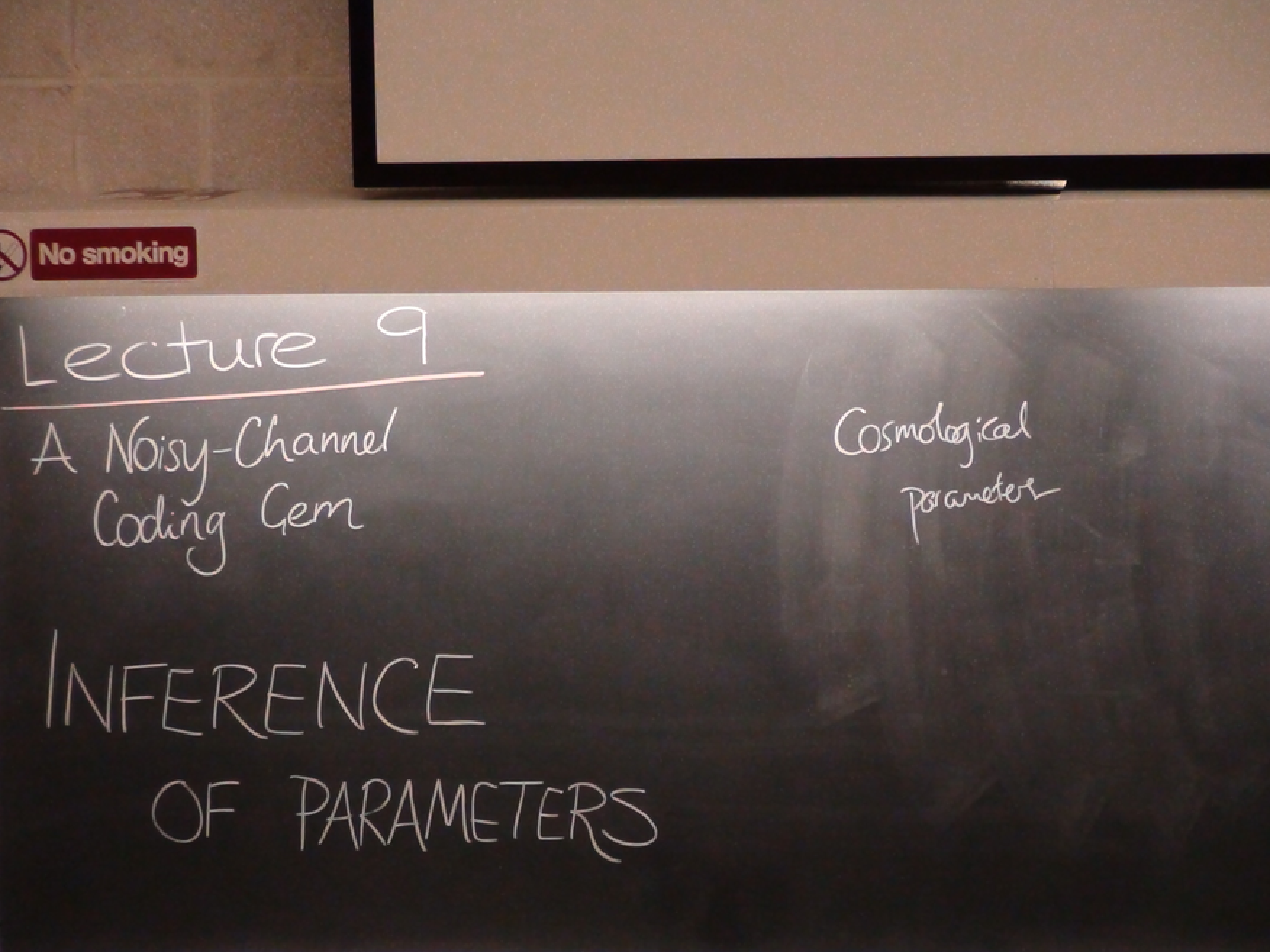
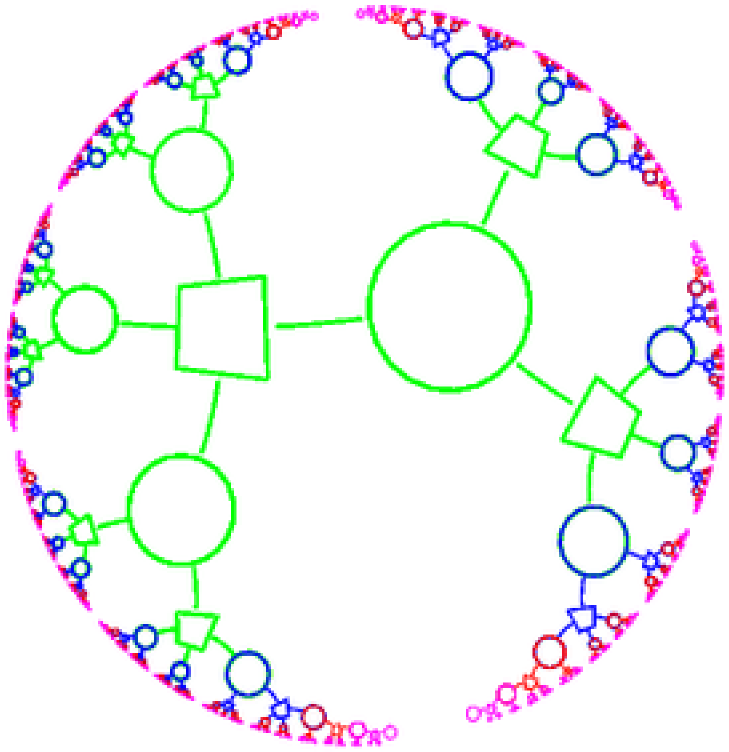

No smoking
Lecture 9
A Noisy-Channel Coding Gem
Cosmological parameters
INFERENCE OF PARAMETERS



Information theory, pattern recognition, and neural networks

- 1 Noisy-channel coding

Source coding (Data compression)

- 2 Information content, entropy
- 3 Typicality and the source coding theorem
- 4 Symbol codes
- 5 Symbol codes and Arithmetic coding

Noisy-channel coding

- 6 Inference and
  Information measures for noisy channels
- 7 Capacity of a noisy channel
- 8 The noisy-channel coding theorem

Inference

- 9 Inference of parameters

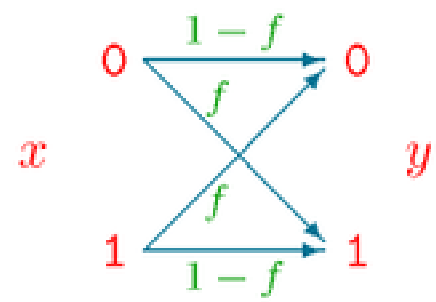
Capacity

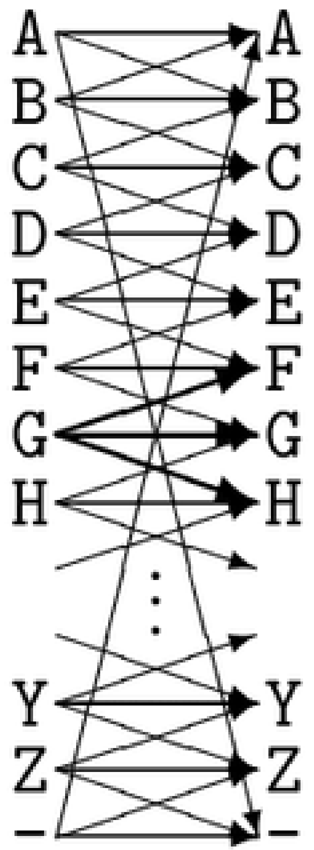
The Capacity of a channel

is the maximum, over all input distributions $P(x)$, of the mutual information:

$$C \equiv \max_{\mathcal{P}_X} I(X;Y)$$

The distribution $\mathcal{P}_X^*$ that achieves the maximum is called the optimal input distribution.



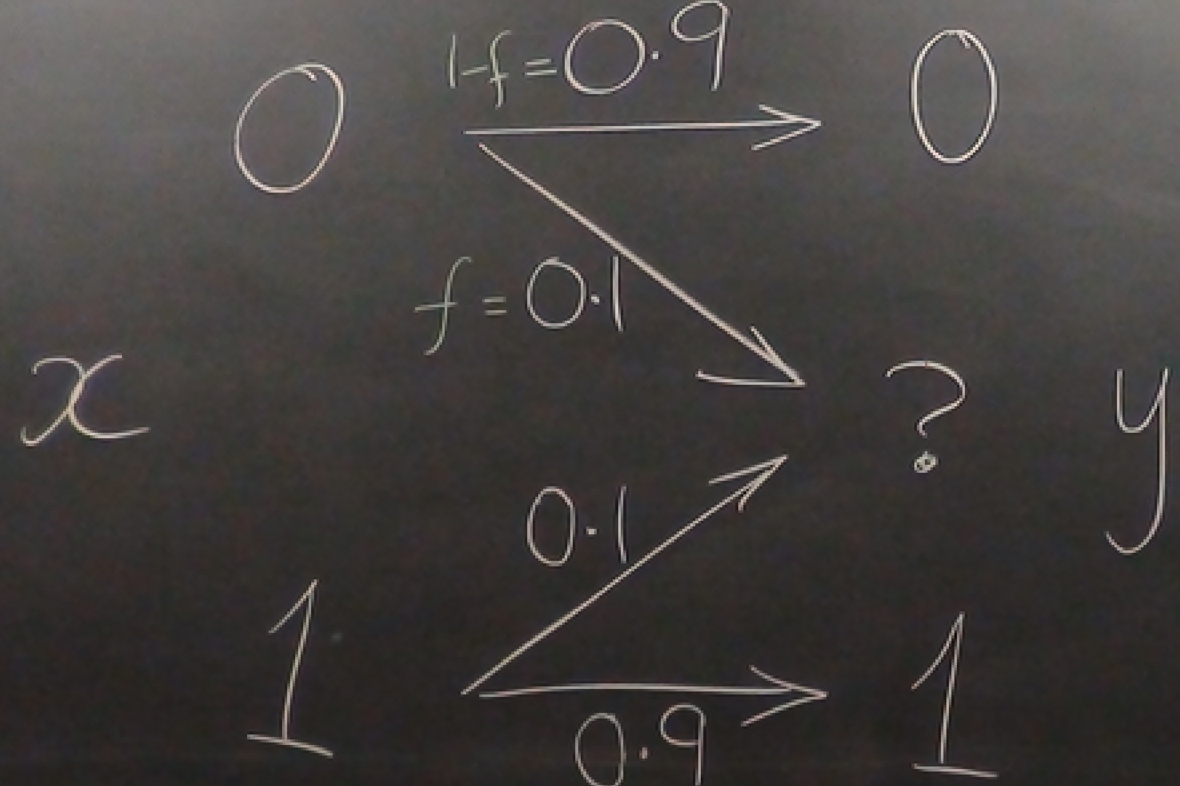


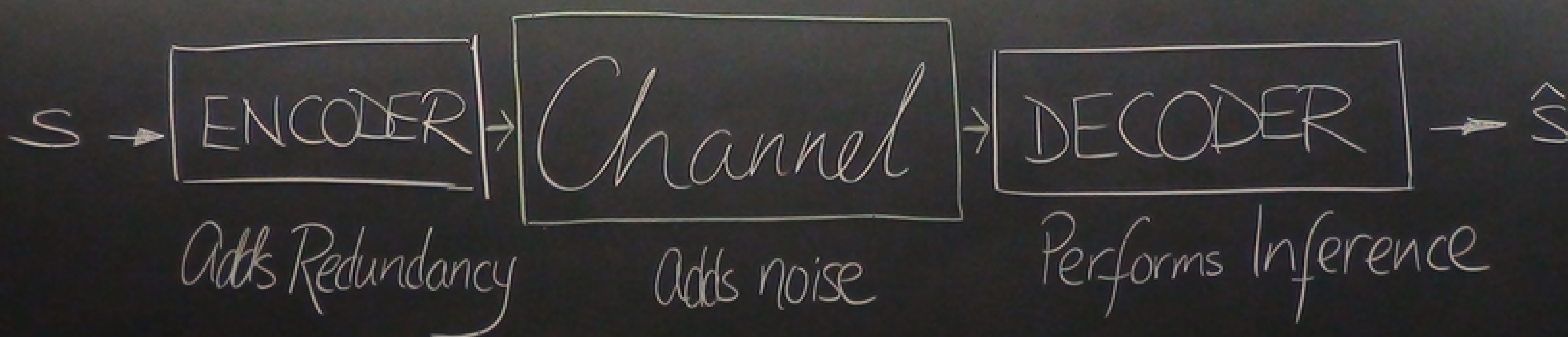
- Shannon's noisy channel coding theorem:

Reliable (virtually error-free) communication is possible at rates up to C

$$I(X;Y) = H(X) - H(X|Y)$$

What is the capacity of the binary




S
ENCODER
Channel
DECODER
Ŝ
Adds Redundancy
Adds noise
Performs Inference

$$I(X;Y) = H(X) - H(X|Y)$$

$$= H_2(p_0) - f\ H_2(p_0)$$

$$= H_2(p_0)\ (1-f)$$

$$= 1\ .\ (1-f)$$

$$= H(Y) - H(Y|X)$$

$$H\big(p(y=0),\ p(y=?),\ p(y=1)\big) - \overset{\checkmark}{H_2(f)}$$

$p(y=0)$ ↑ $p_0 \times (1-f)$

Decomposability of entropy

$$H(Y) \overset{?}{=} H_2(f) + (1-f) \times H_2(p_0)$$

if not, ...

$= ?), P(y=1) \Big)$ $-$ $H_2(f)$ ✓

...y of entropy

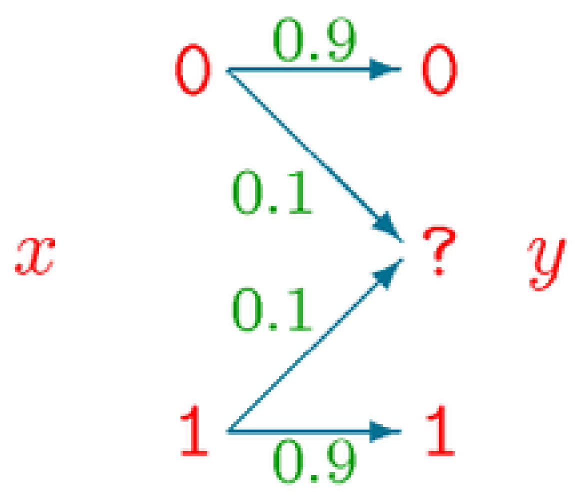
$(1-f) \times H_2(p_0)$

$I(X;Y)$

$= \cancel{H_2(f)} + (1-f) H_2(p_0)$

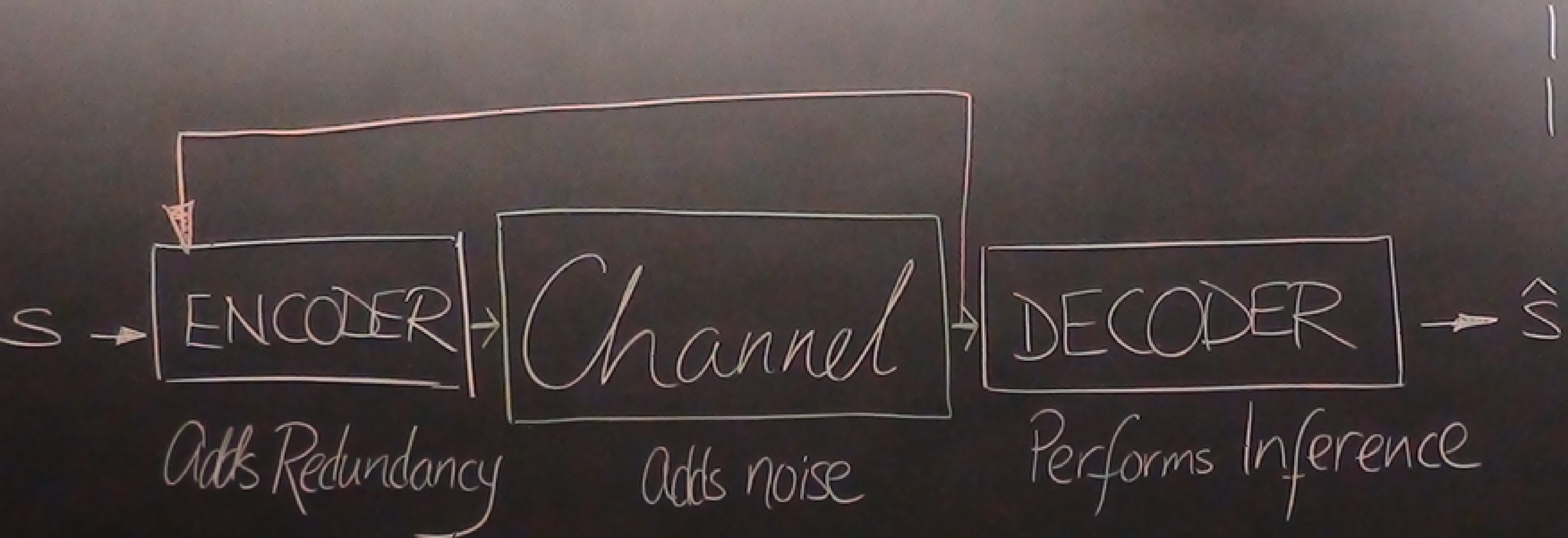
$- \cancel{H_2(f)}$

Binary erasure channel



$$\mathbf{Q} = \begin{bmatrix} 0.9 & 0 \\ 0.1 & 0.1 \\ 0 & 0.9 \end{bmatrix}$$

What is the capacity?


s
ENCODER
Channel
DECODER
ŝ
Adds Redundancy
Adds noise
Performs Inference

1 0 0 1 0 1 1 1

1 0 0 ? 0 1 1 ?

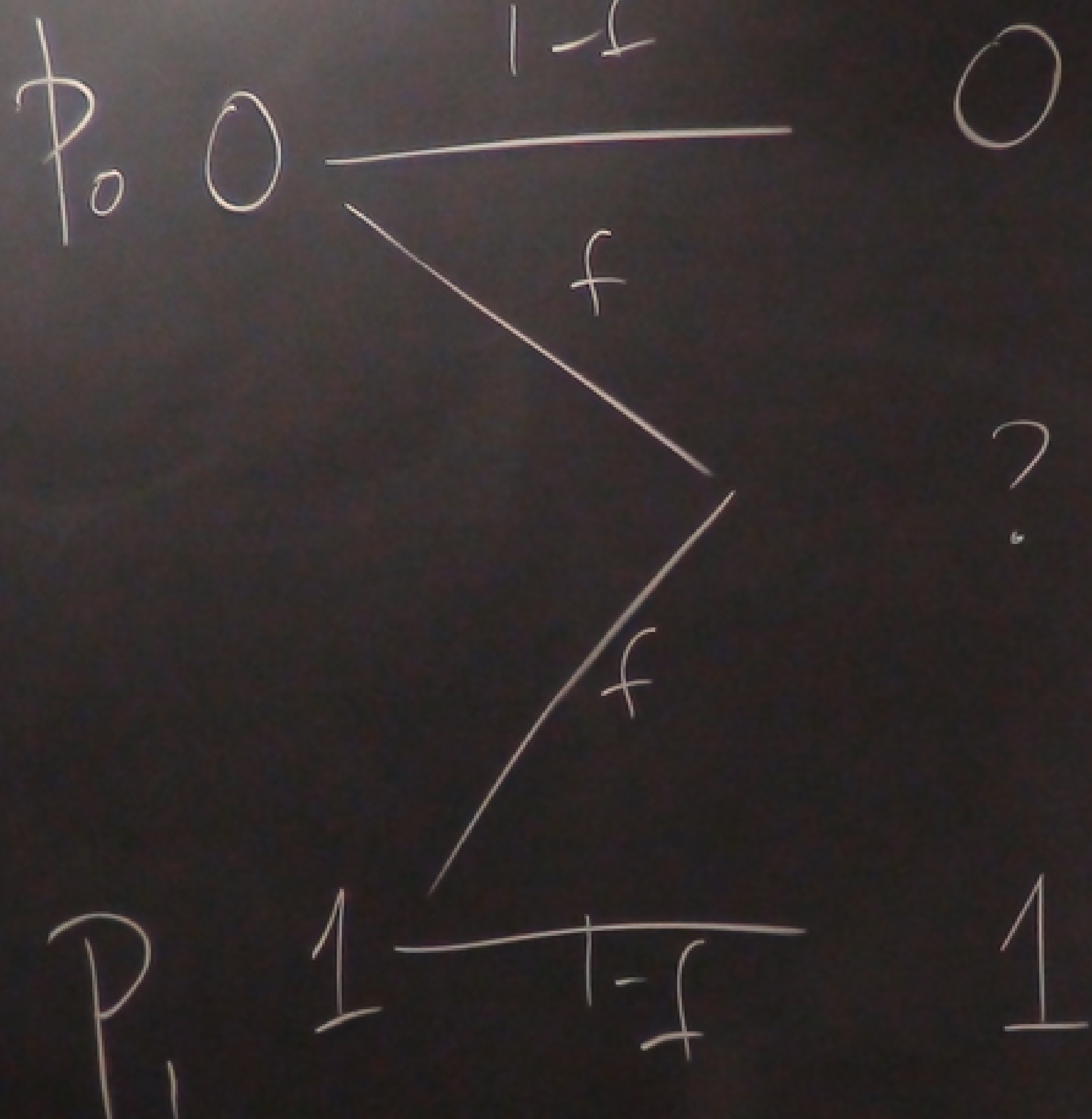
source 1 0 0 1 0 1 1 1

? ?

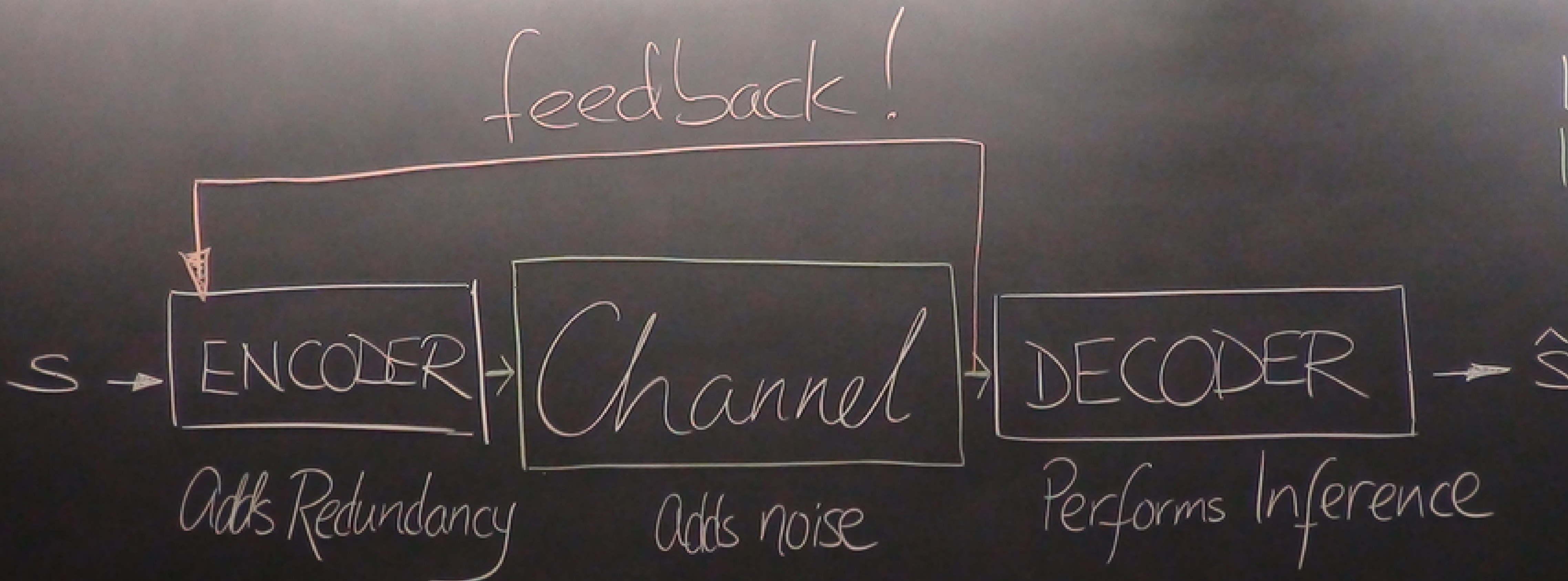
DER → $\hat{S}$ transmissio 1 0 0 1 1 0 1 1 1

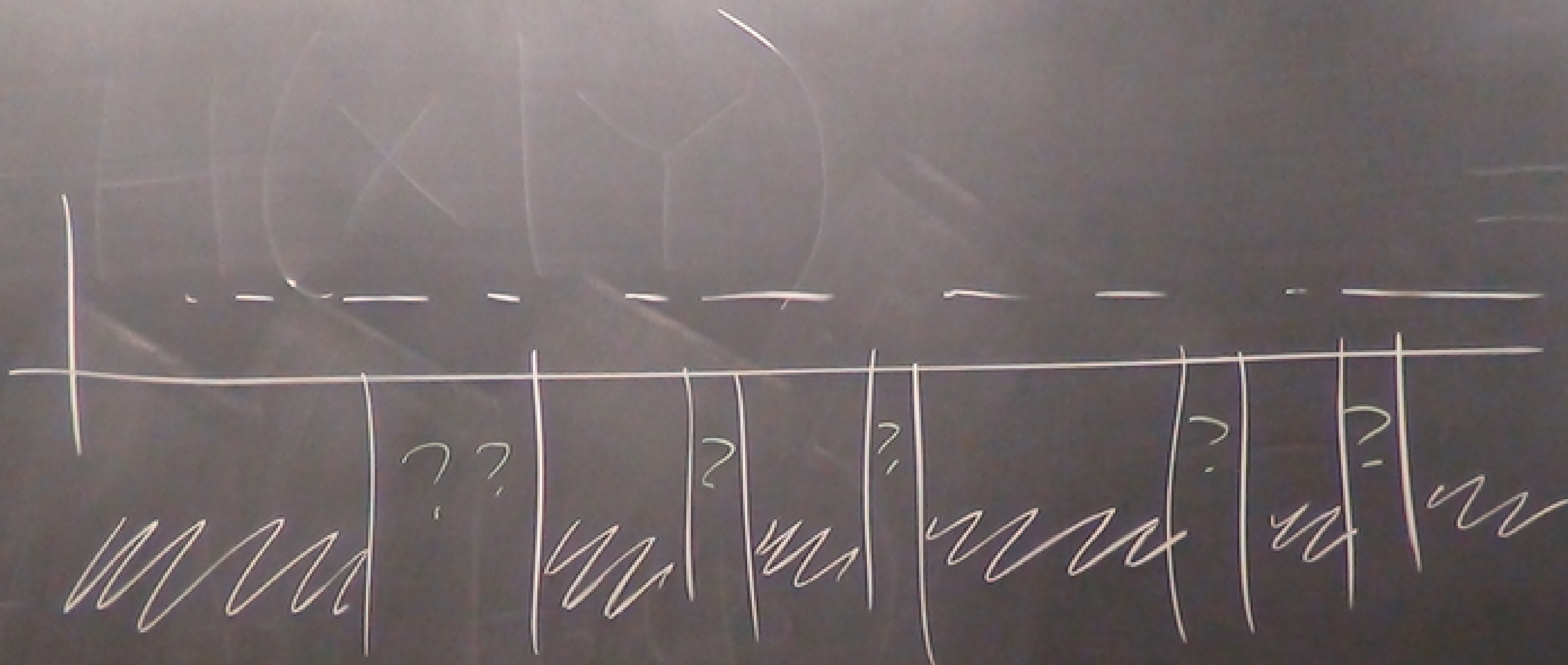
s Inference

Achievable Rate $= 1 - f$

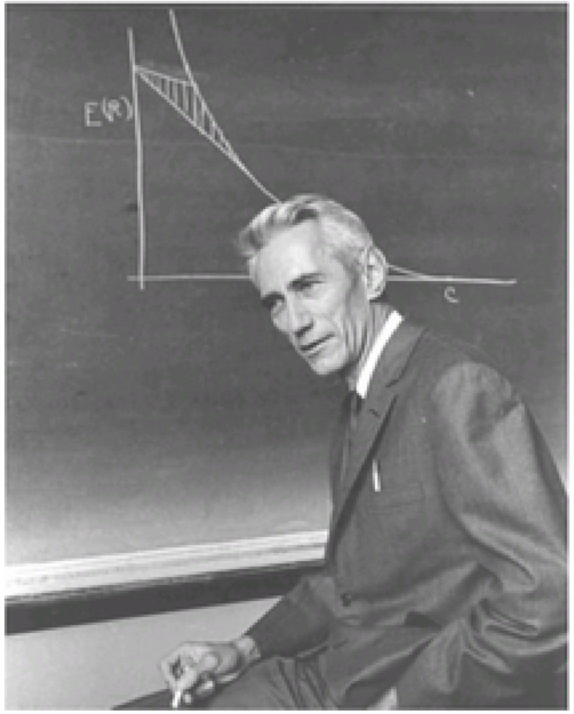


p_0 0
1-f
0
f
?
f
P_1 1
1-f
1


feedback!
S
ENCODER
Channel
DECODER
Ŝ
Adds Redundancy
Adds noise
Performs Inference



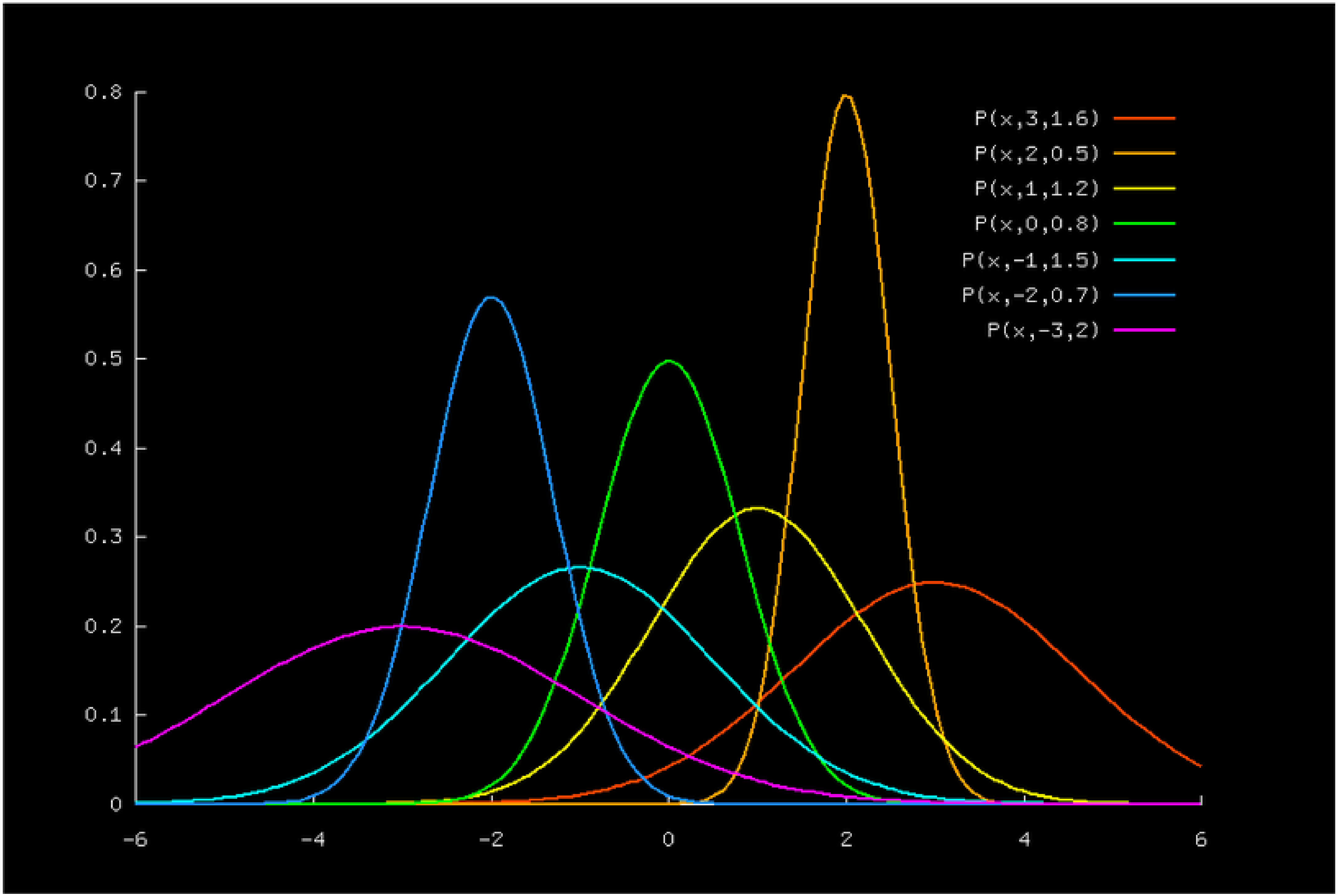
Feedback

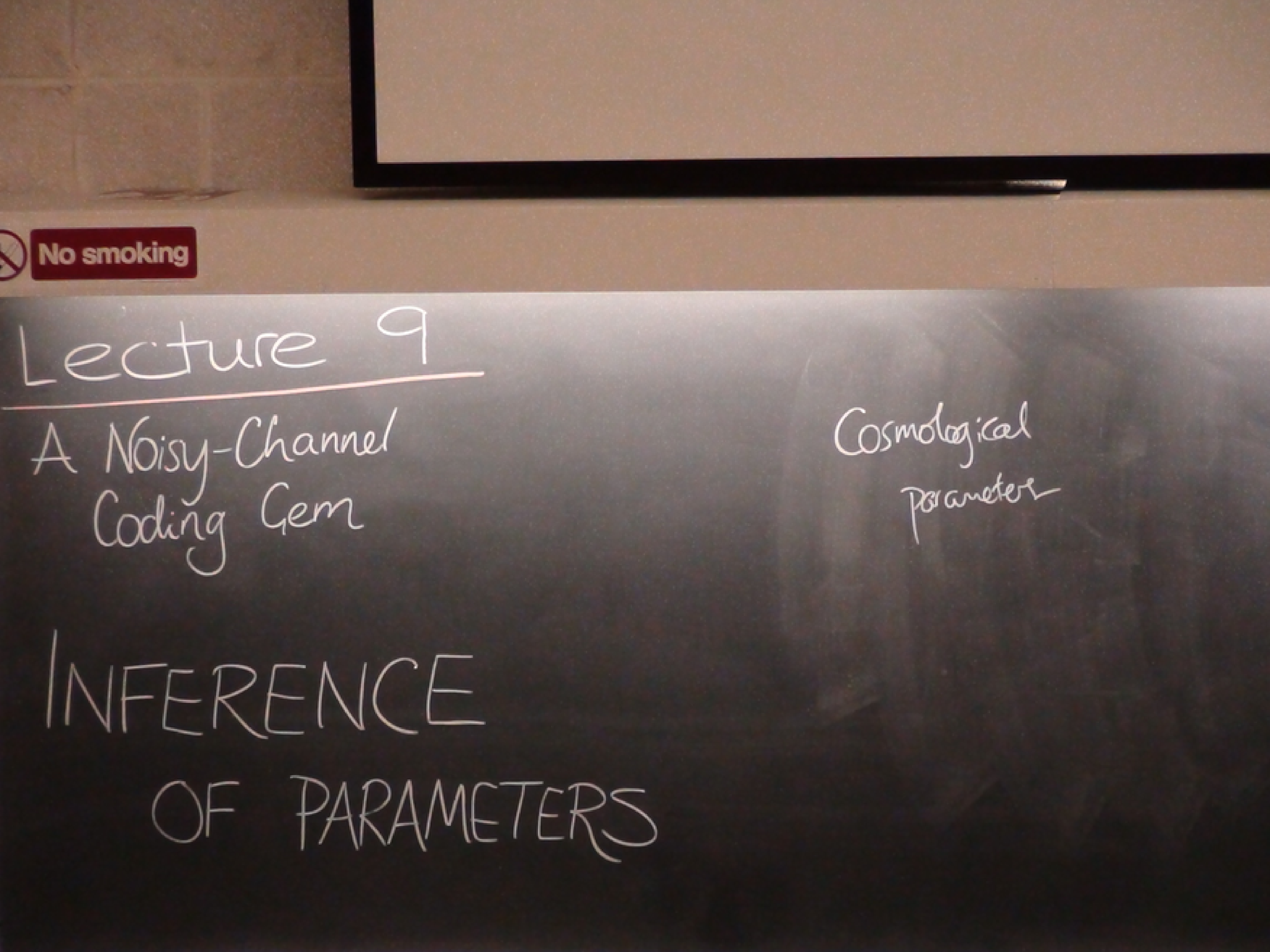


`Feedback? Pah! Who needs feedback?

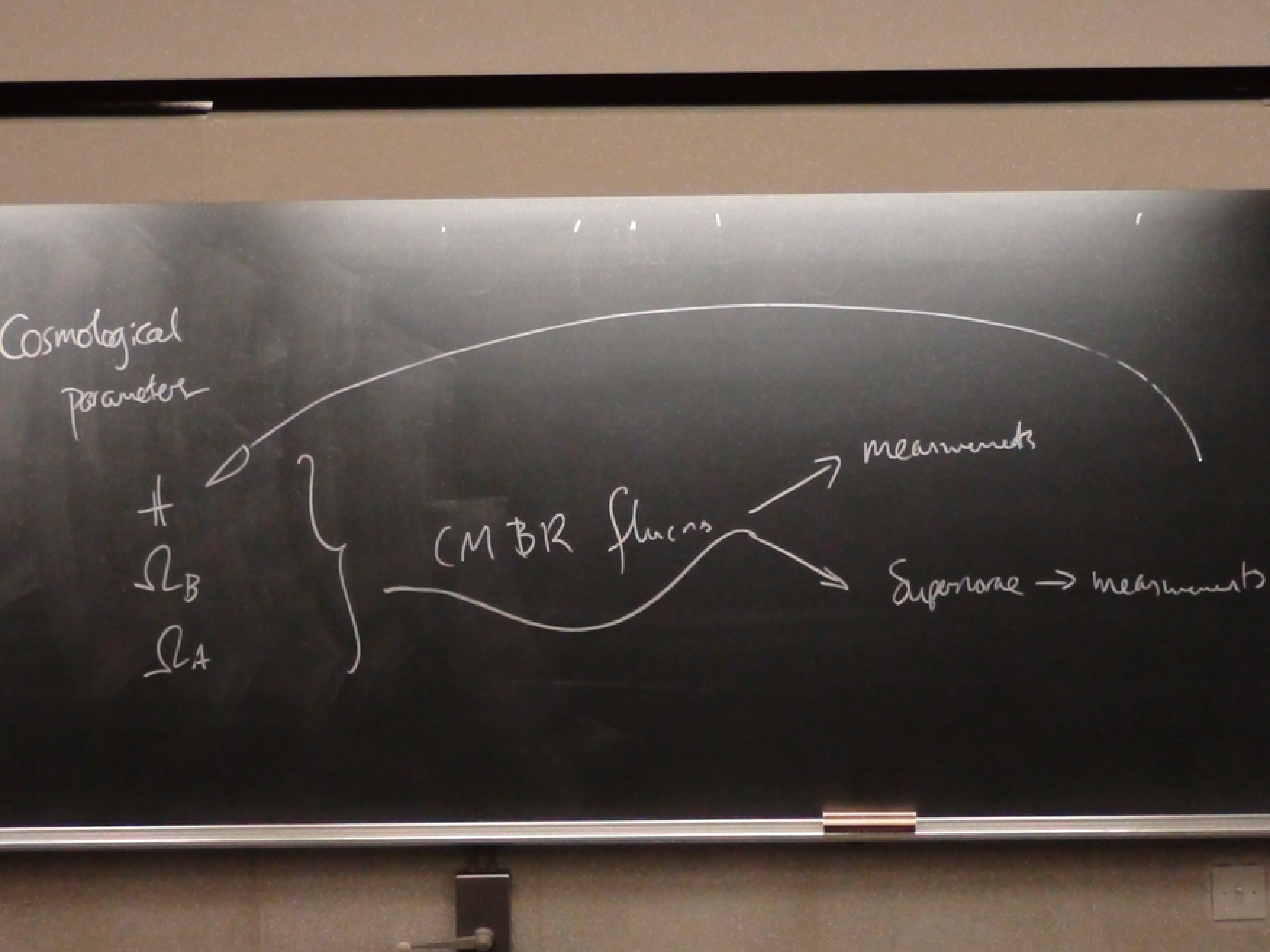
Just use a random code!'

Inferring the parameters of a Gaussian





No smoking
Lecture 9
A Noisy-Channel
Coding Gem
Cosmological
parameters
INFERENCE
OF PARAMETERS


Cosmological parameters
H
ΩB
ΩΛ
CMBR flucns
measurements
Supernovae → measurements

Whether spam

$$s_1 \longrightarrow e_1$$

$$s_2 \longrightarrow e_2$$

Image rec

Speech rec

Character rec


email
space

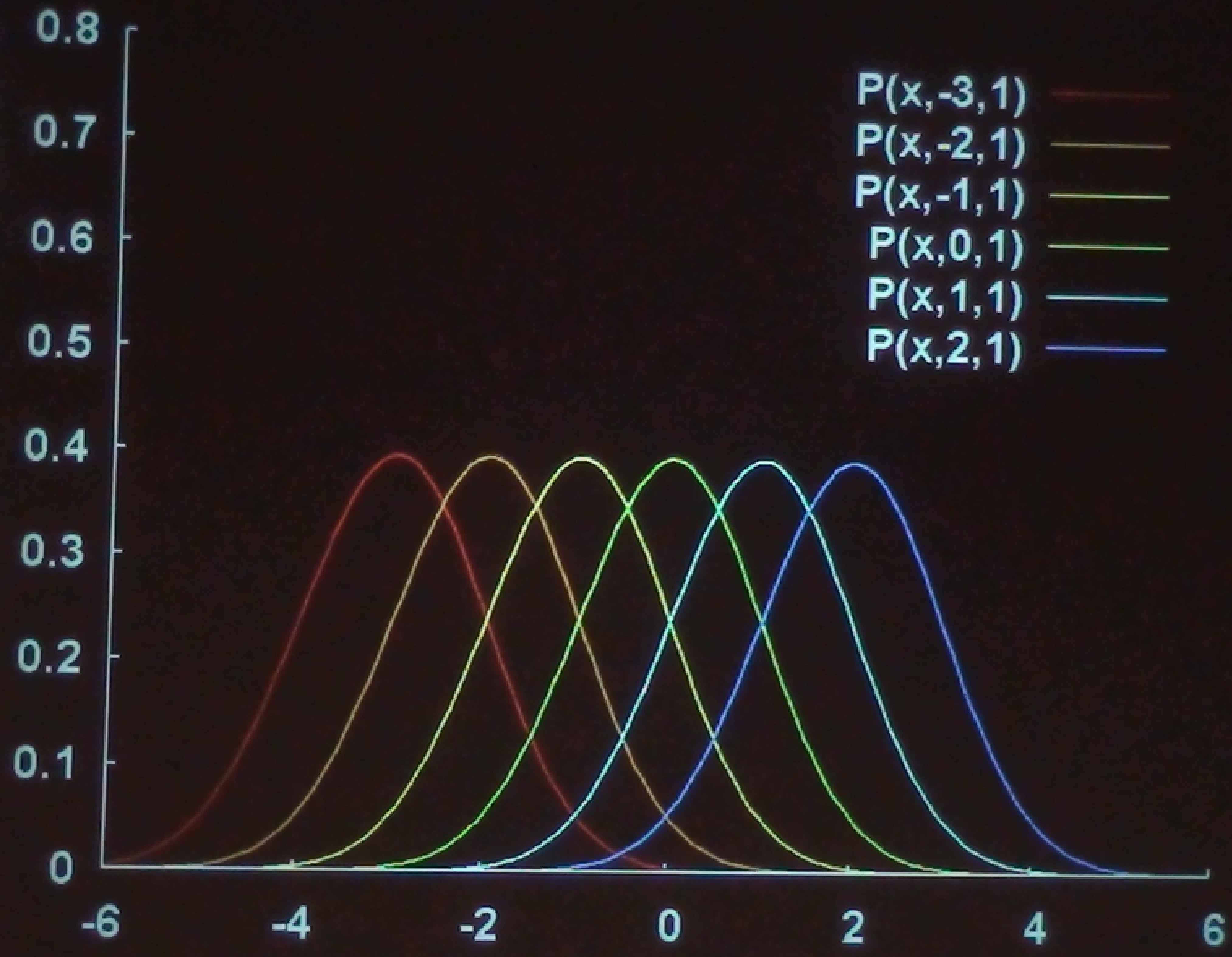
$$\mu, \sigma \longrightarrow \{ x_n \}_{n=1}^{N}$$

N

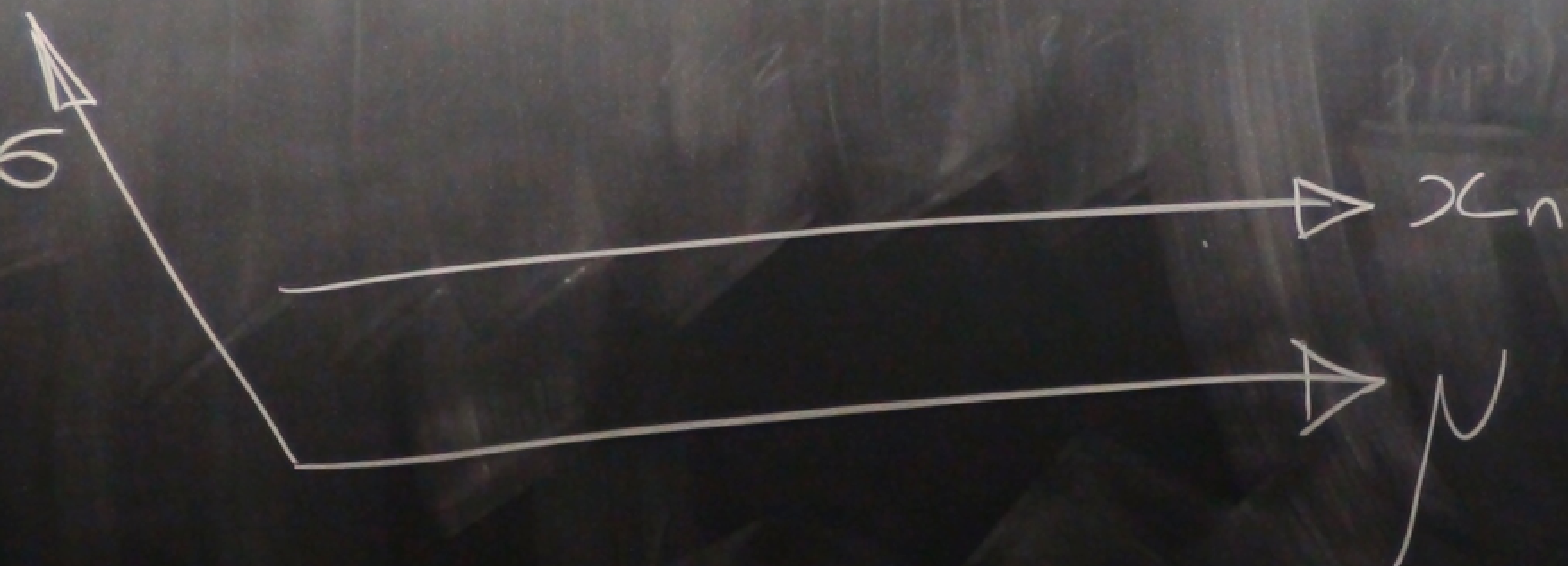
$n = 1 \ldots N$

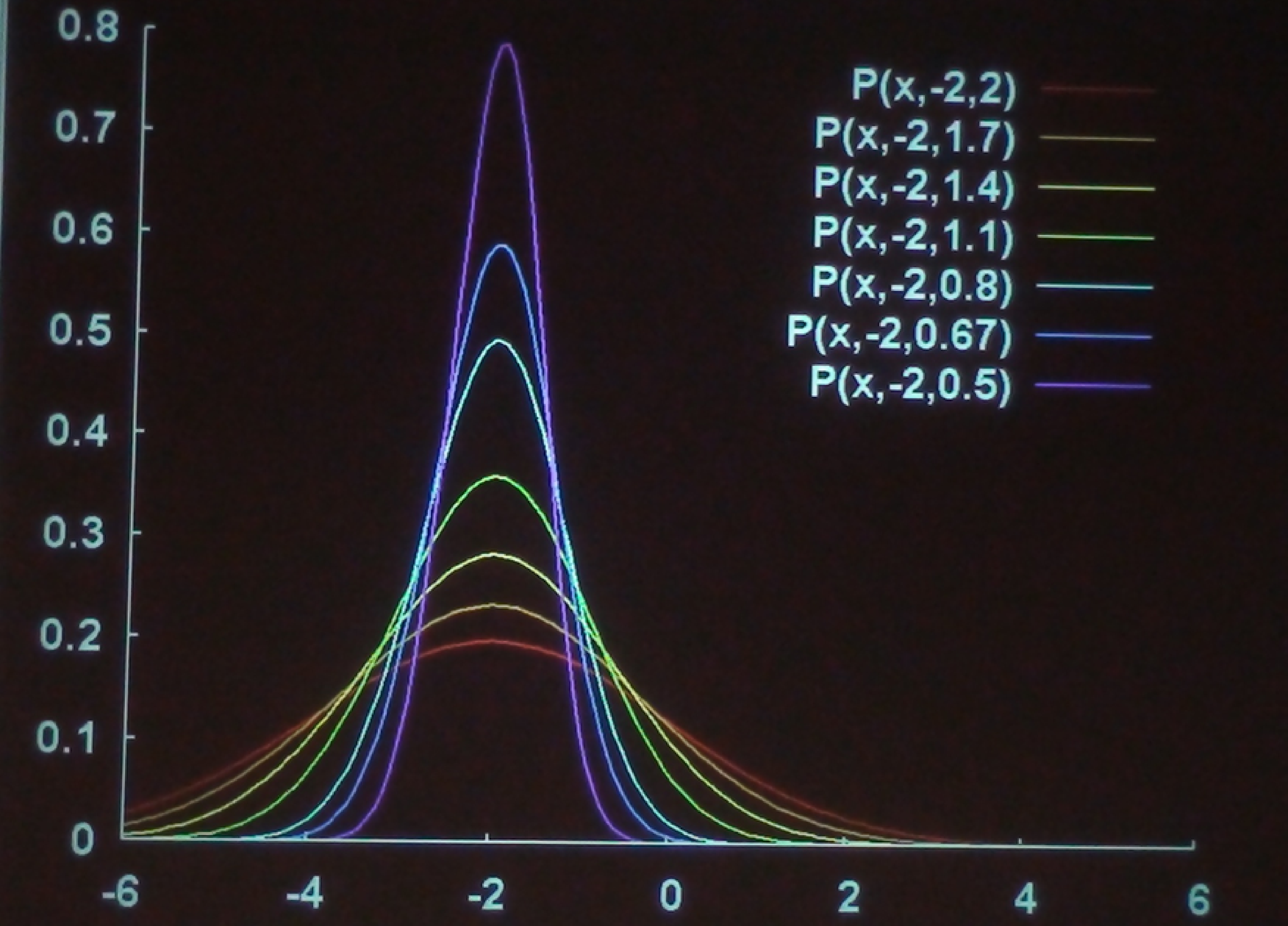
$$P(x_n \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_n - \mu)^2}{2\sigma^2}}$$

Infer μ, σ given $\{x\}$


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-6
-4
-2
0
2
4
6
P(x,-3,1)
P(x,-2,1)
P(x,-1,1)
P(x,0,1)
P(x,1,1)
P(x,2,1)

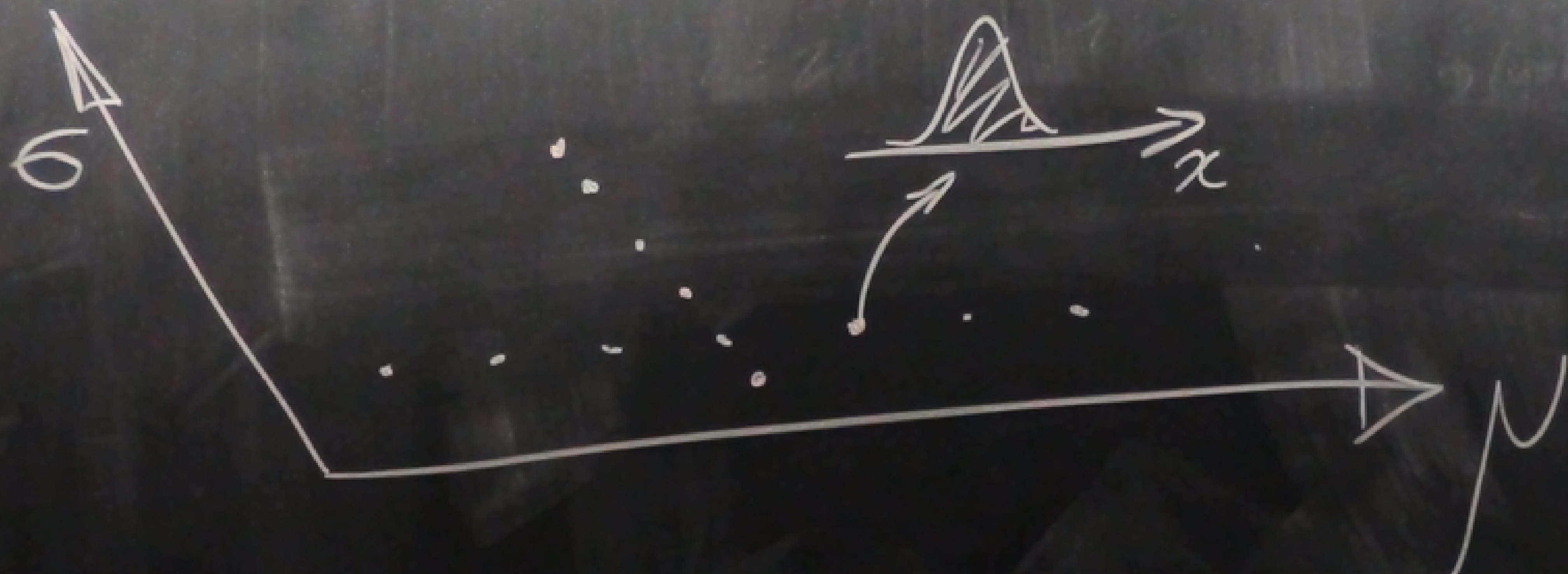
Infer μ, σ given $\{x\}$

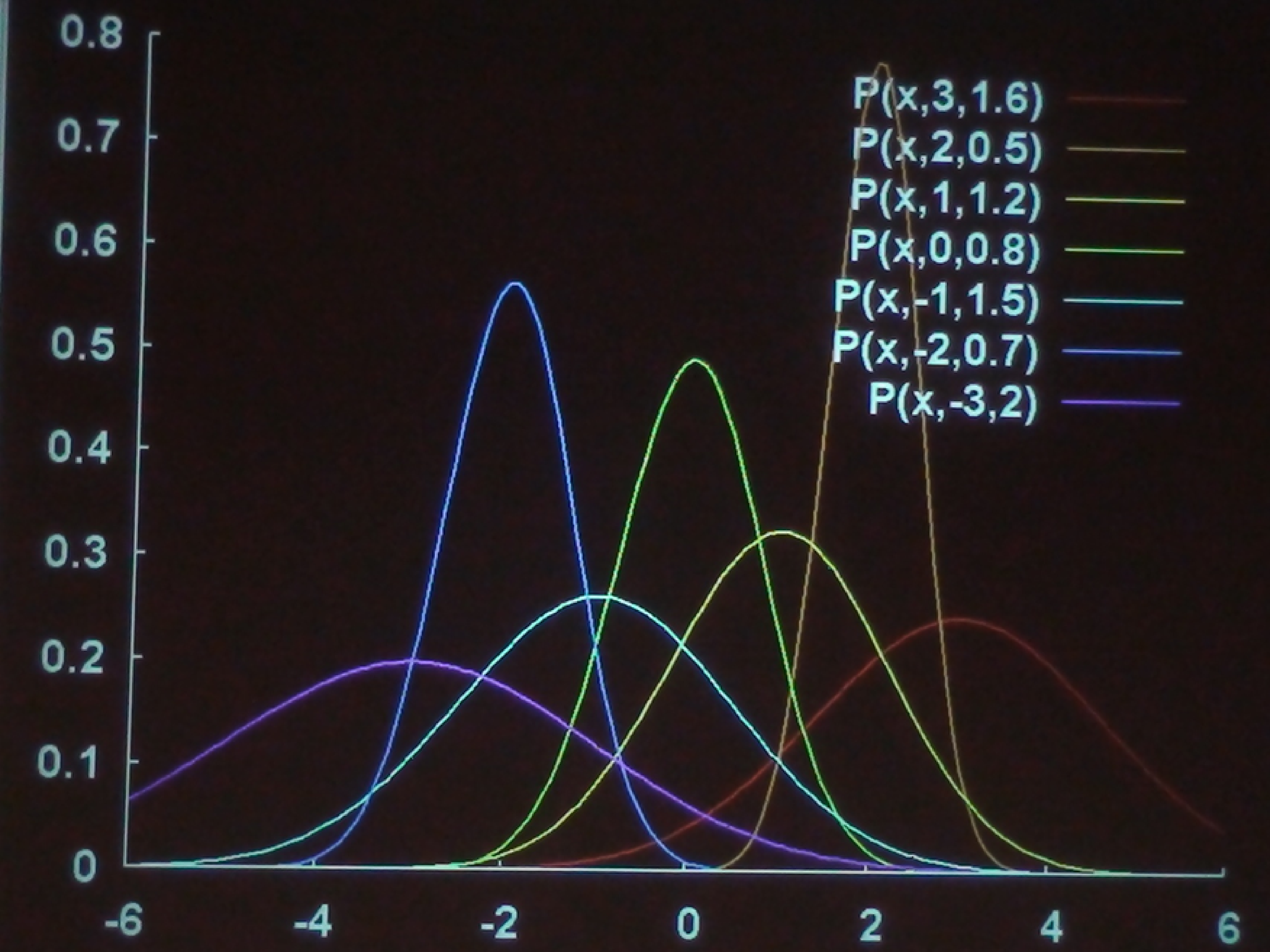



0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-6
-4
-2
0
2
4
6
P(x,-2,2)
P(x,-2,1.7)
P(x,-2,1.4)
P(x,-2,1.1)
P(x,-2,0.8)
P(x,-2,0.67)
P(x,-2,0.5)

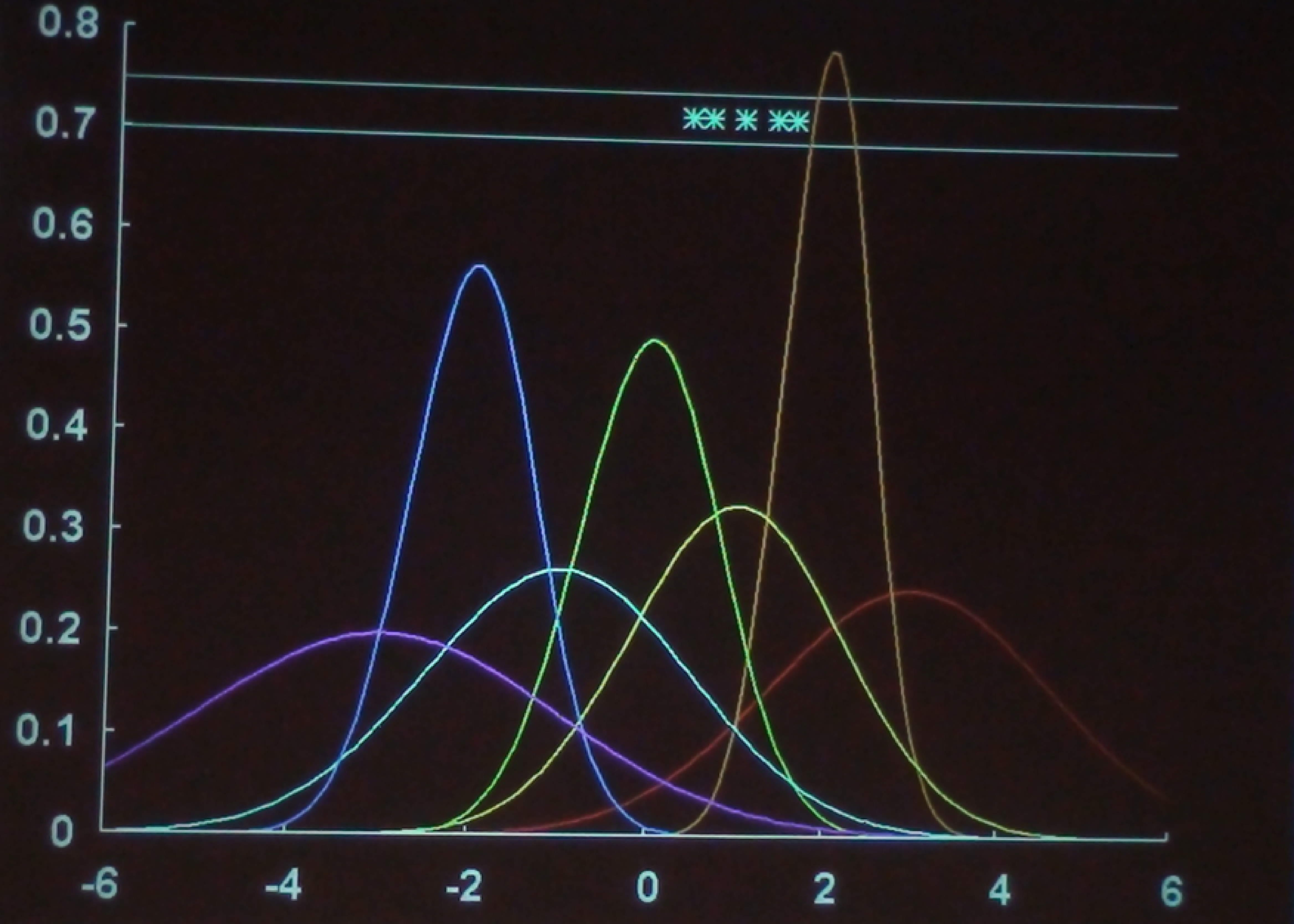
Infer μ, σ given $\{x\}$

N





0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-6
-4
-2
0
2
4
6
P(x,3,1.6)
P(x,2,0.5)
P(x,1,1.2)
P(x,0,0.8)
P(x,-1,1.5)
P(x,-2,0.7)
P(x,-3,2)


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-6
-4
-2
0
2
4
6

$$P(\mu, \sigma \mid \{x_n\}) = \frac{P(\{x \ldots}{P(\ldots}$$

$\hookrightarrow P(\{ \ldots$

$$\frac{P(\{x_n\}|\mu,\sigma)\,P(\mu,\sigma)}{P(\{x_n\})}$$

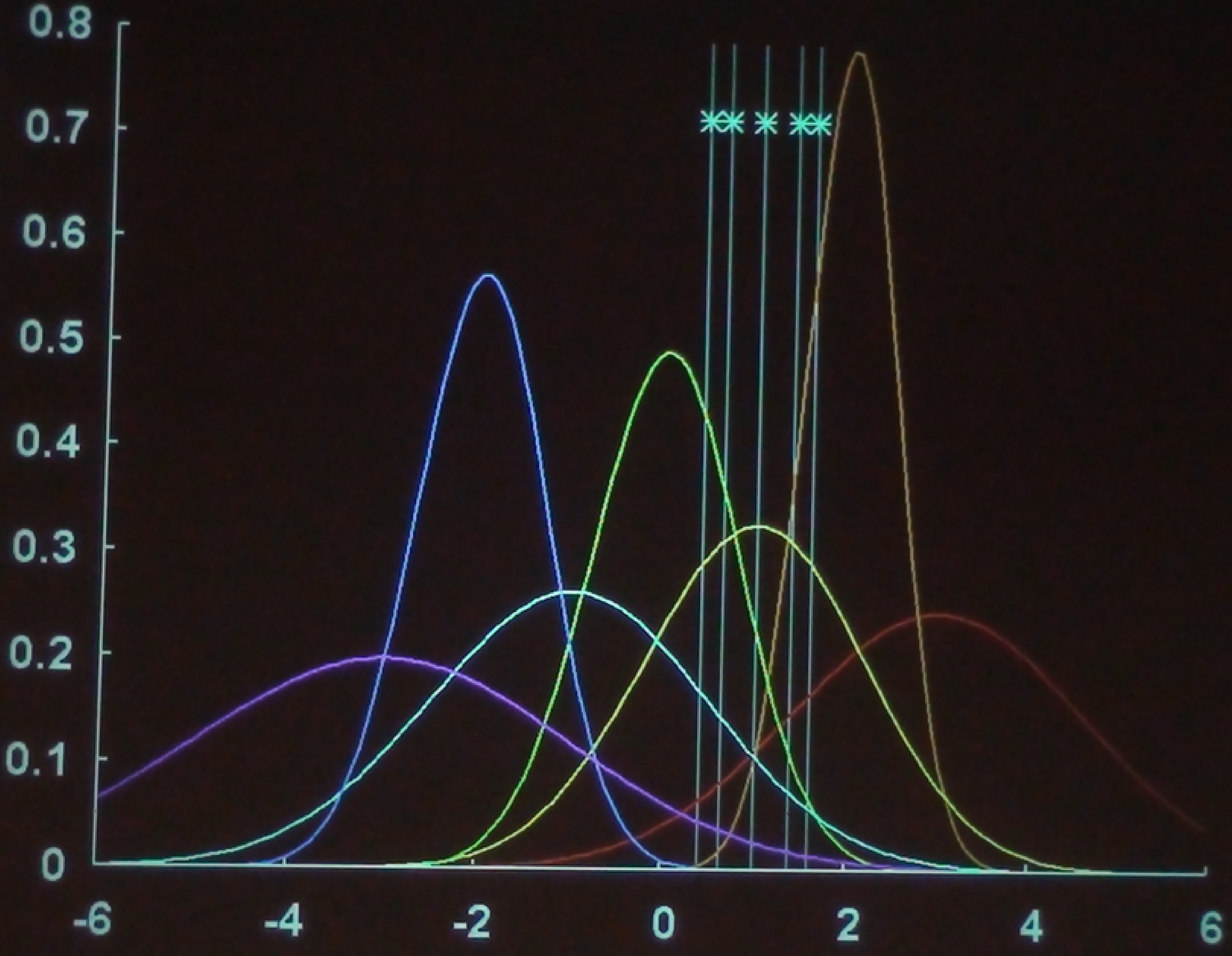
Whether spam

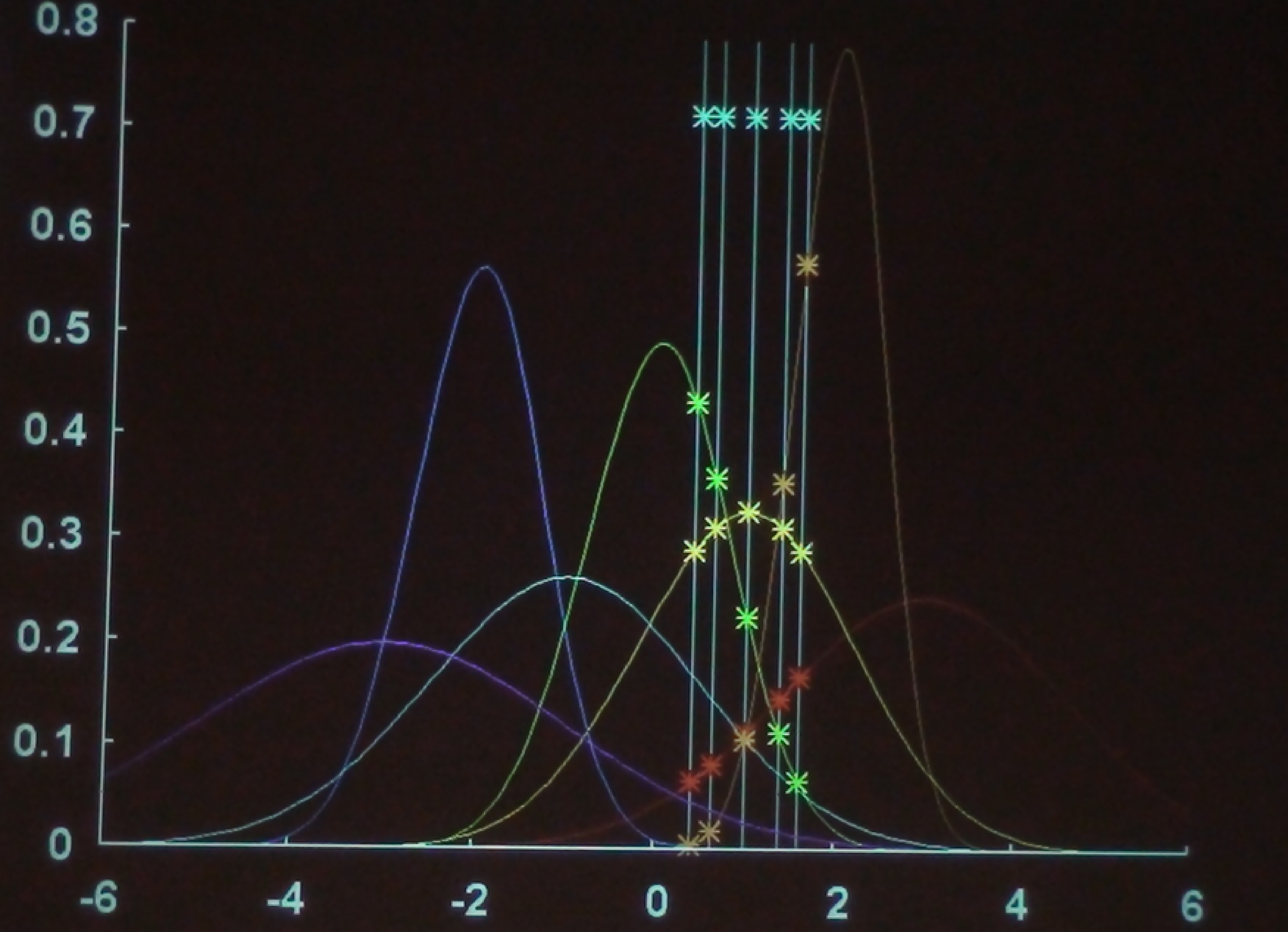
$s_1 \longrightarrow e_1$

$s_2 \longrightarrow e_2$

$\vdots$

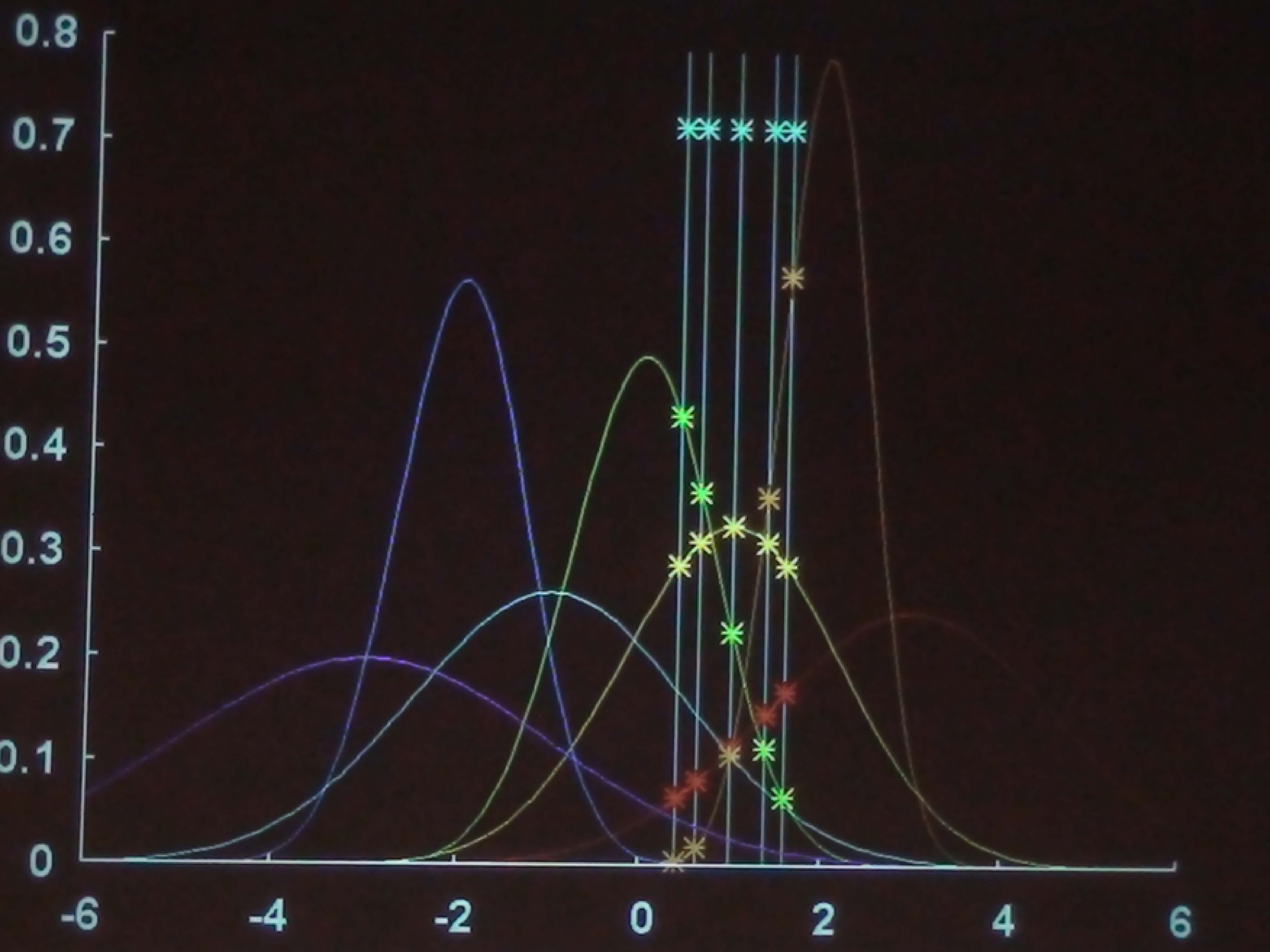
$$P(\{x\}) = \int d\mu\, d\sigma\; P(\{x_n\}|\mu,\sigma)\,P(\mu,\sigma)$$

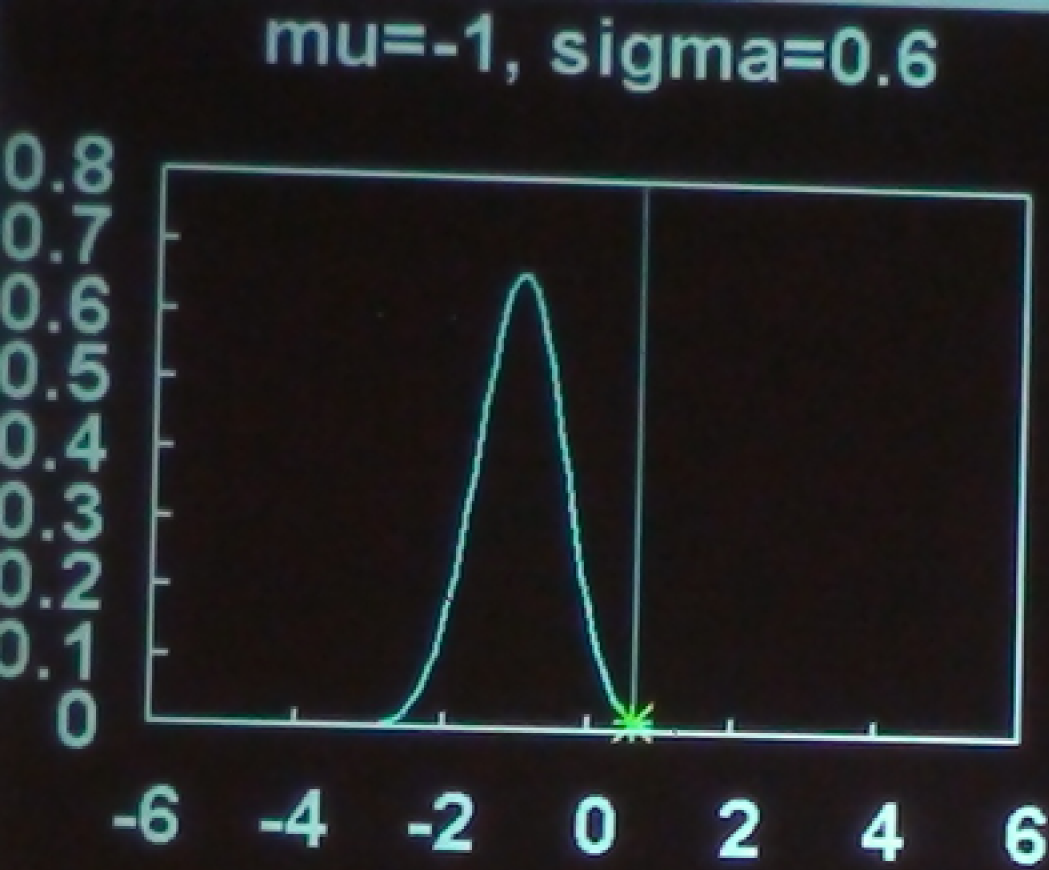

0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-6
-4
-2
0
2
4
6


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-6
-4
-2
0
2
4
6

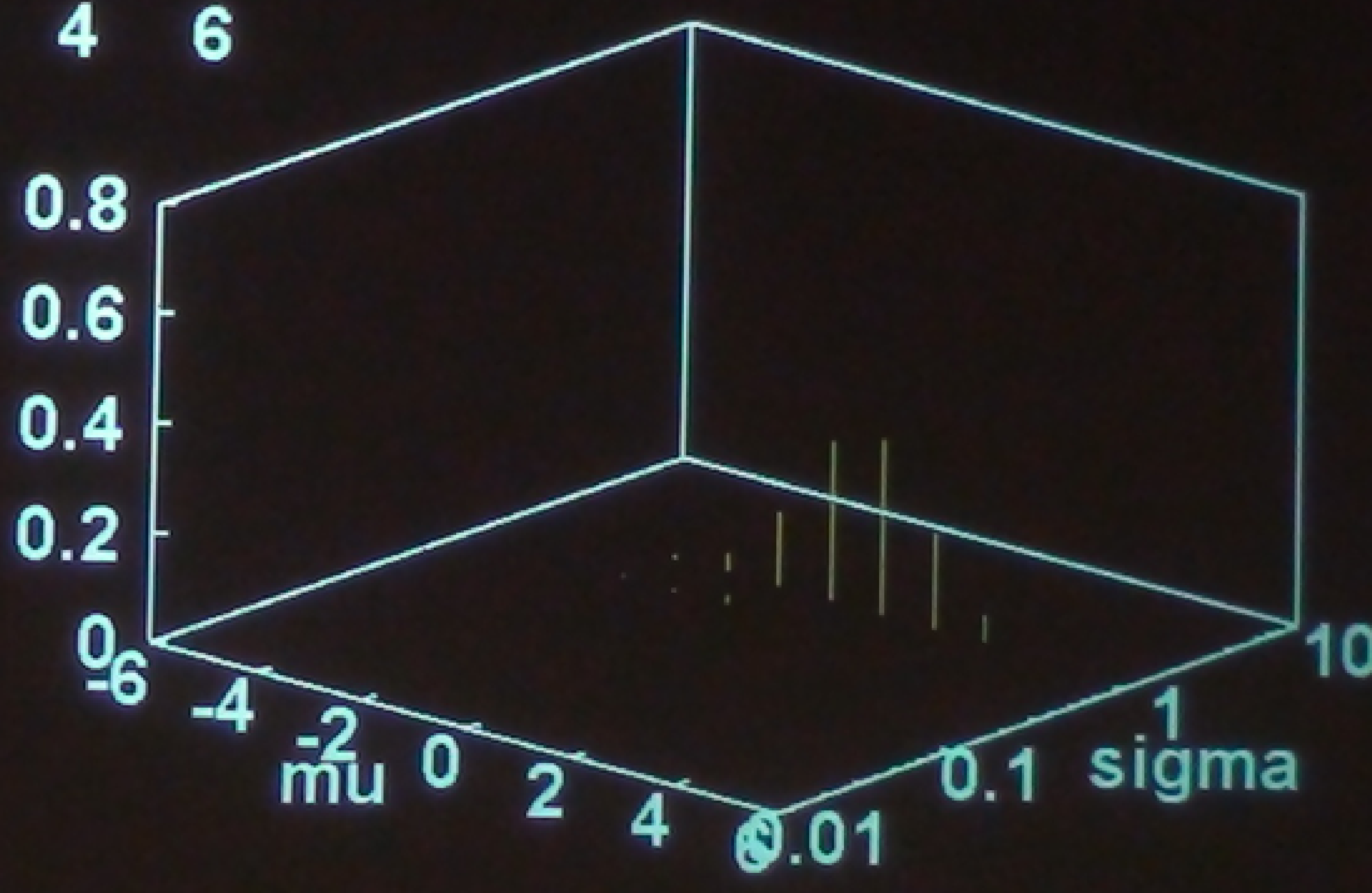
$$P(\mu, \sigma \mid \{x_n\}, \lambda) -$$

$$\prod_{1}^{N} P(x_n \mid \mu, \sigma)$$

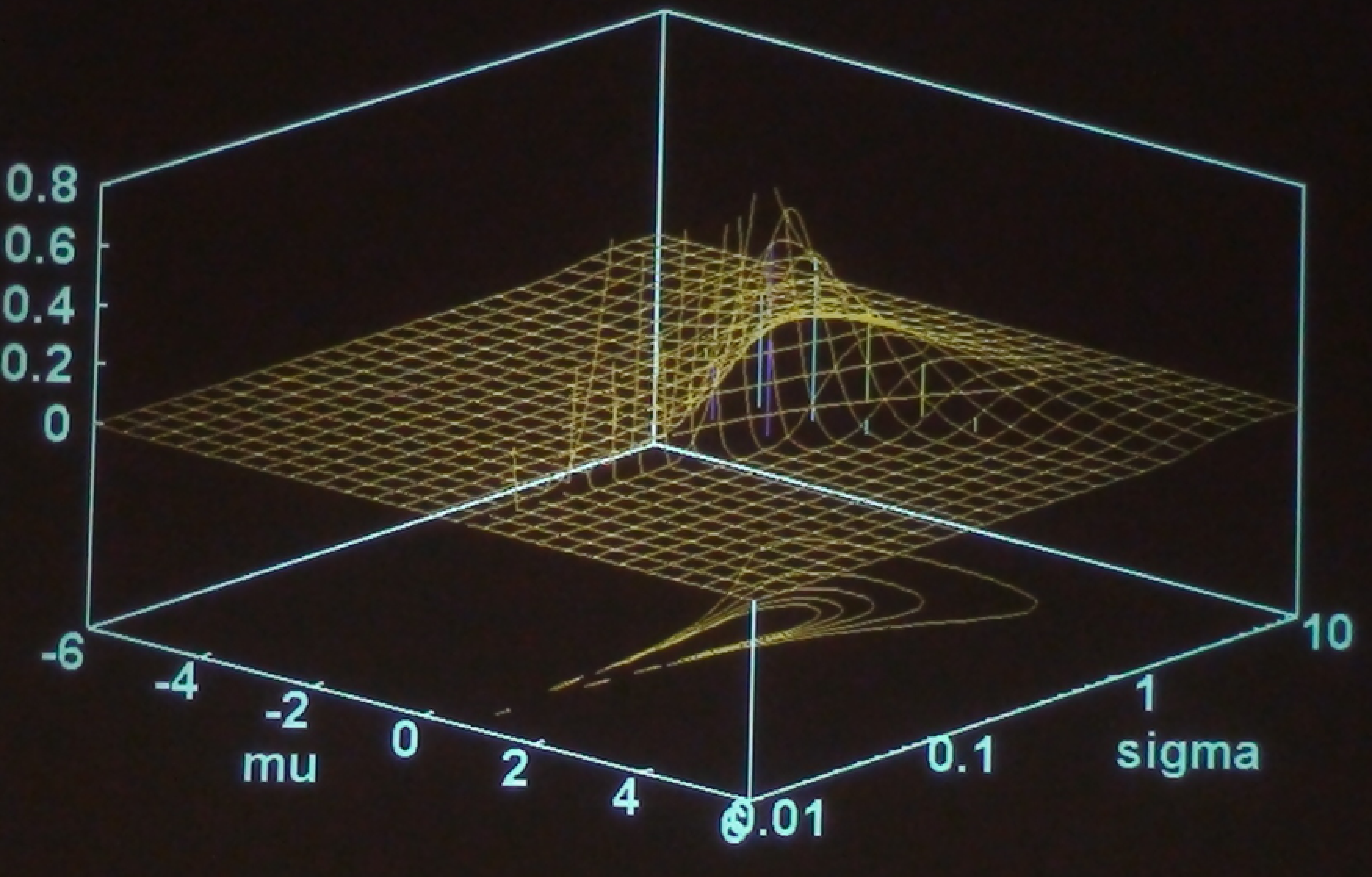

0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-6
-4
-2
0
2
4
6

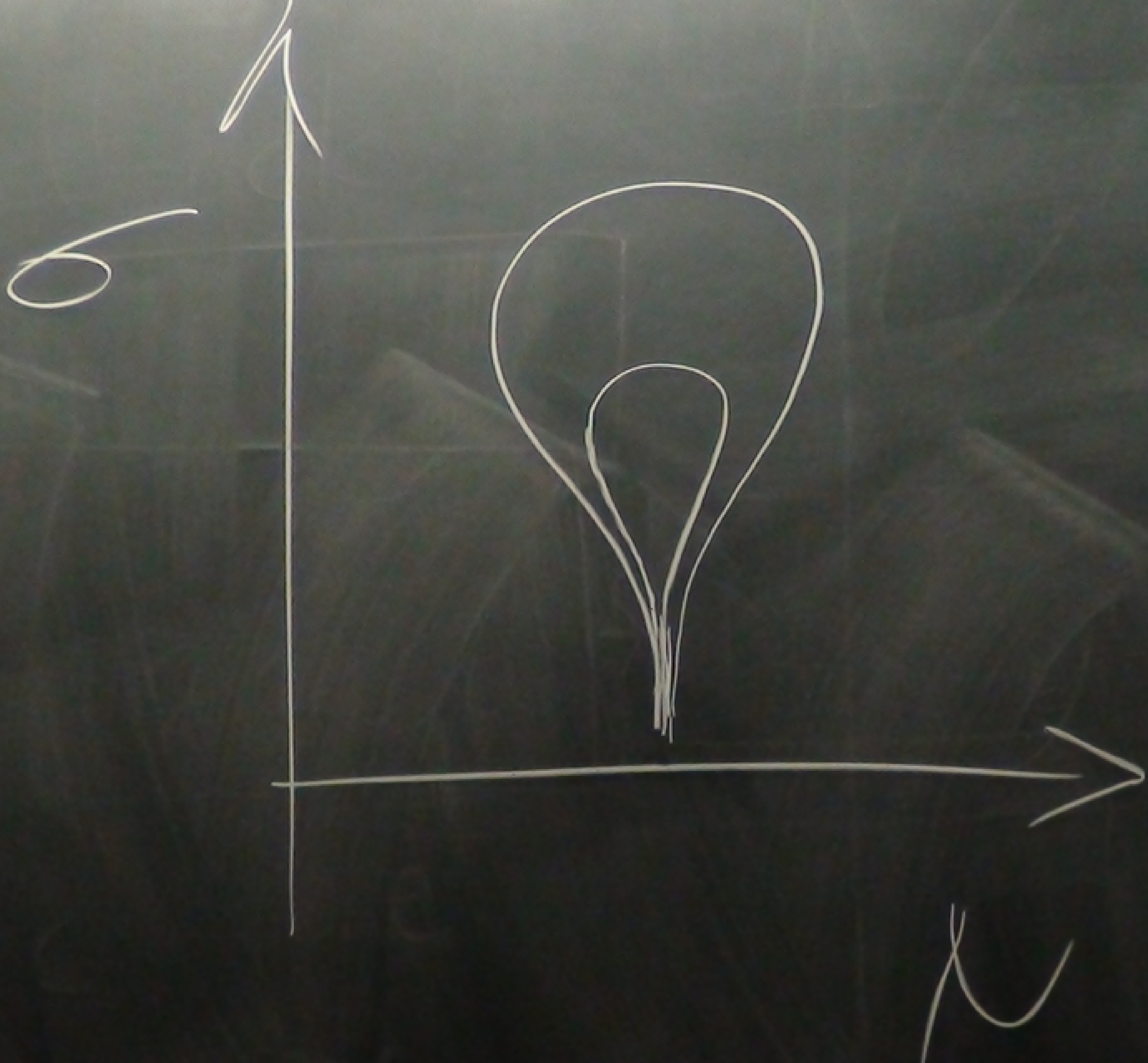


mu=-1, sigma=0.6
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-6
-4
-2
0
2
4
6

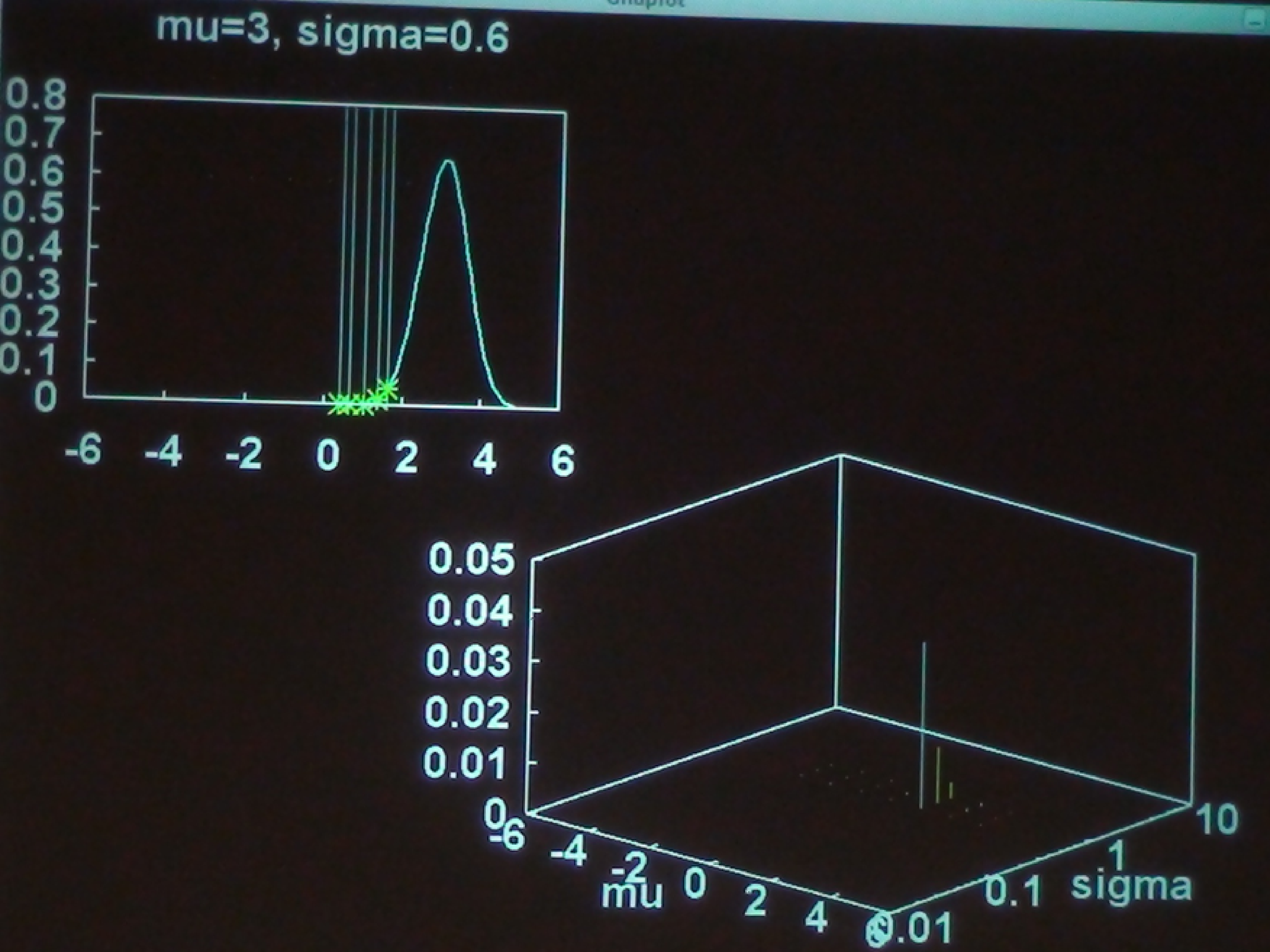


0.8
0.6
0.4
0.2
0
-6
-4
-2
0
2
4
mu
0.01
0.1
1
10
sigma

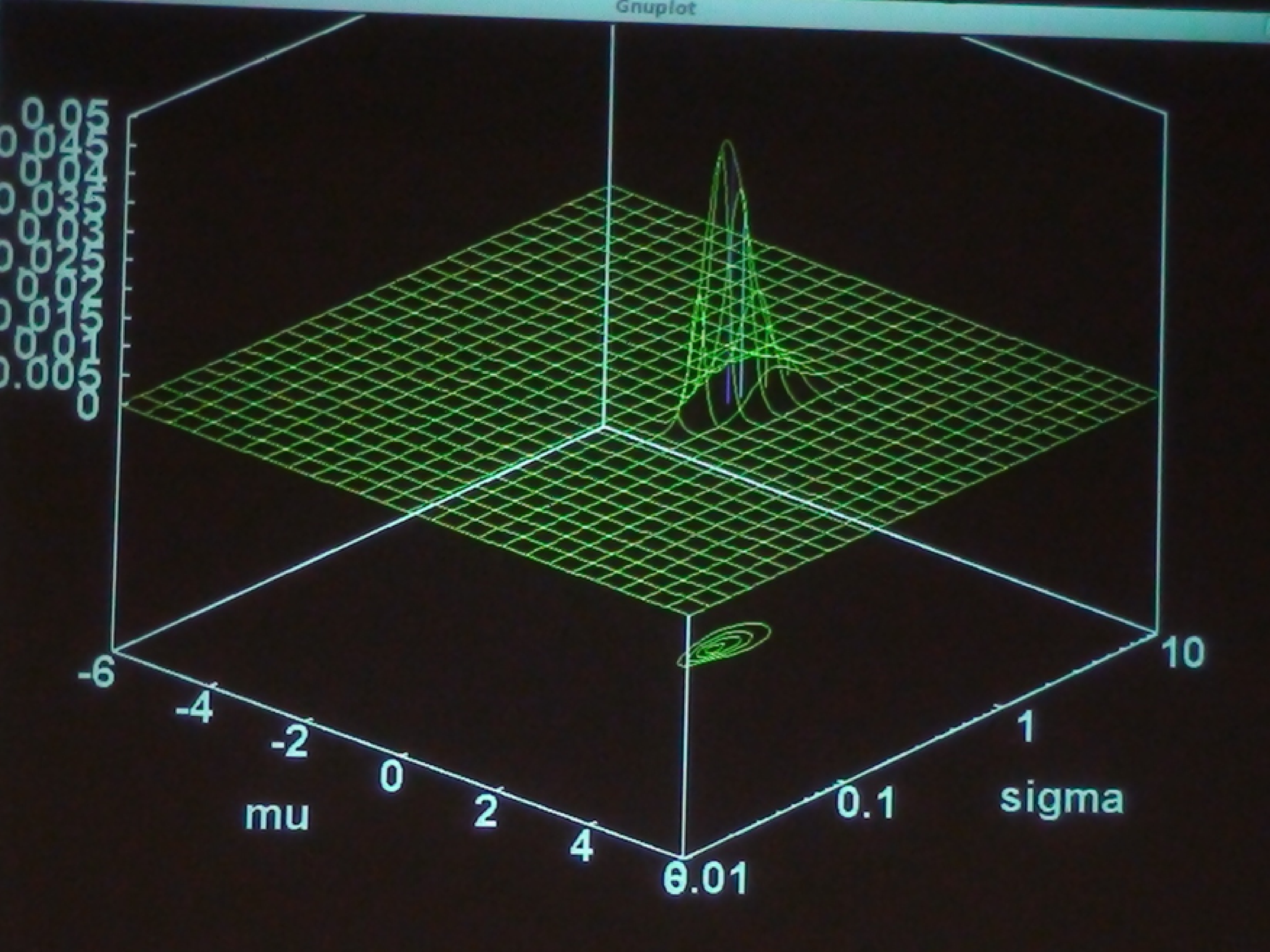

0.8
0.6
0.4
0.2
0
-6
-4
-2
0
2
4
mu
0.01
0.1
1
10
sigma



σ
μ

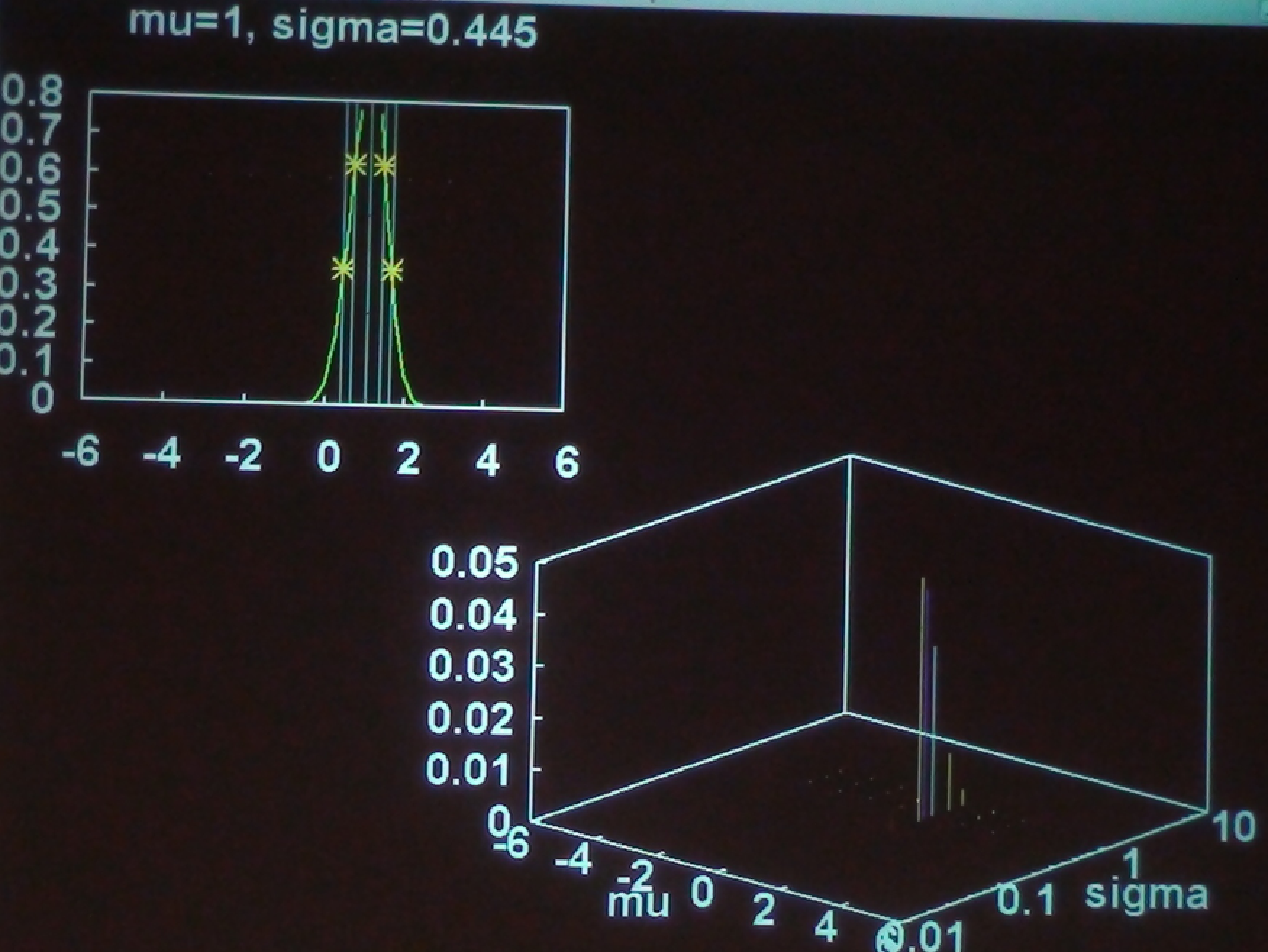


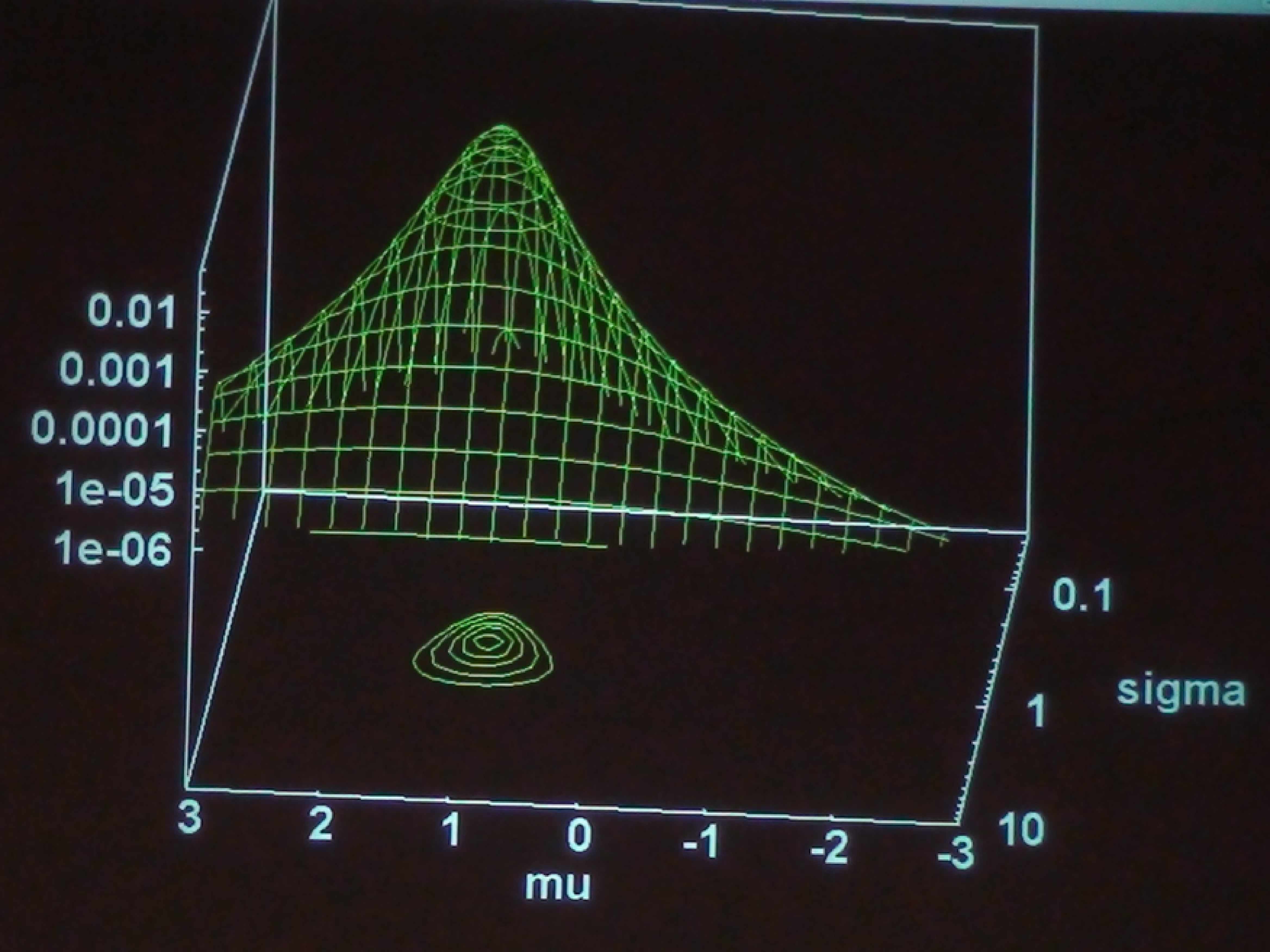
mu=3, sigma=0.6
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-6
-4
-2
0
2
4
6
0.05
0.04
0.03
0.02
0.01
0
-6
-4
-2
0
2
4
6
mu
0.01
0.1
1
10
sigma

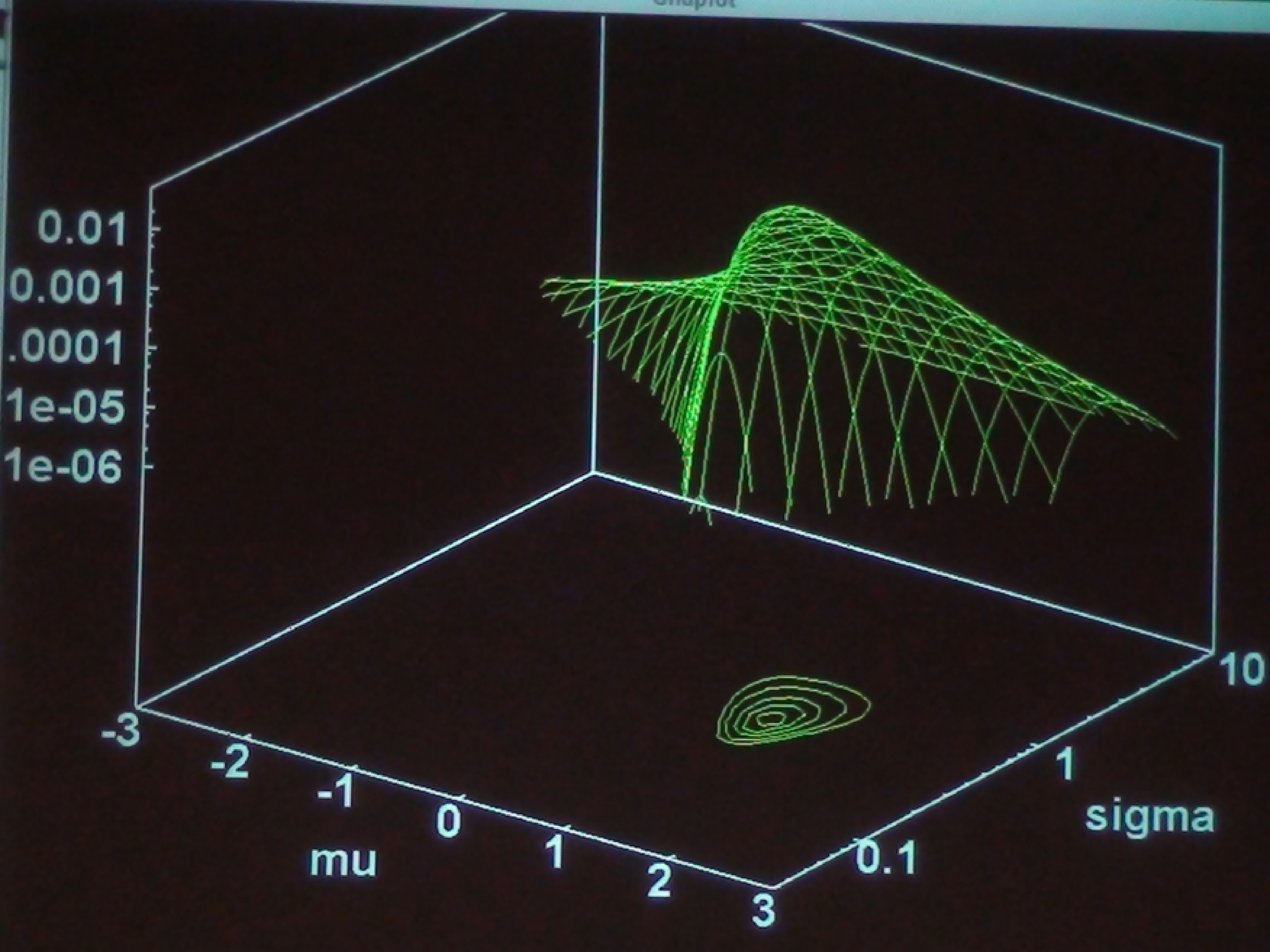

Gnuplot
0.05
0
-6
-4
-2
0
2
4
6
mu
0.01
0.1
1
10
sigma

$N = 5$

$P(\{x_n\} \mid \mu, \sigma)$

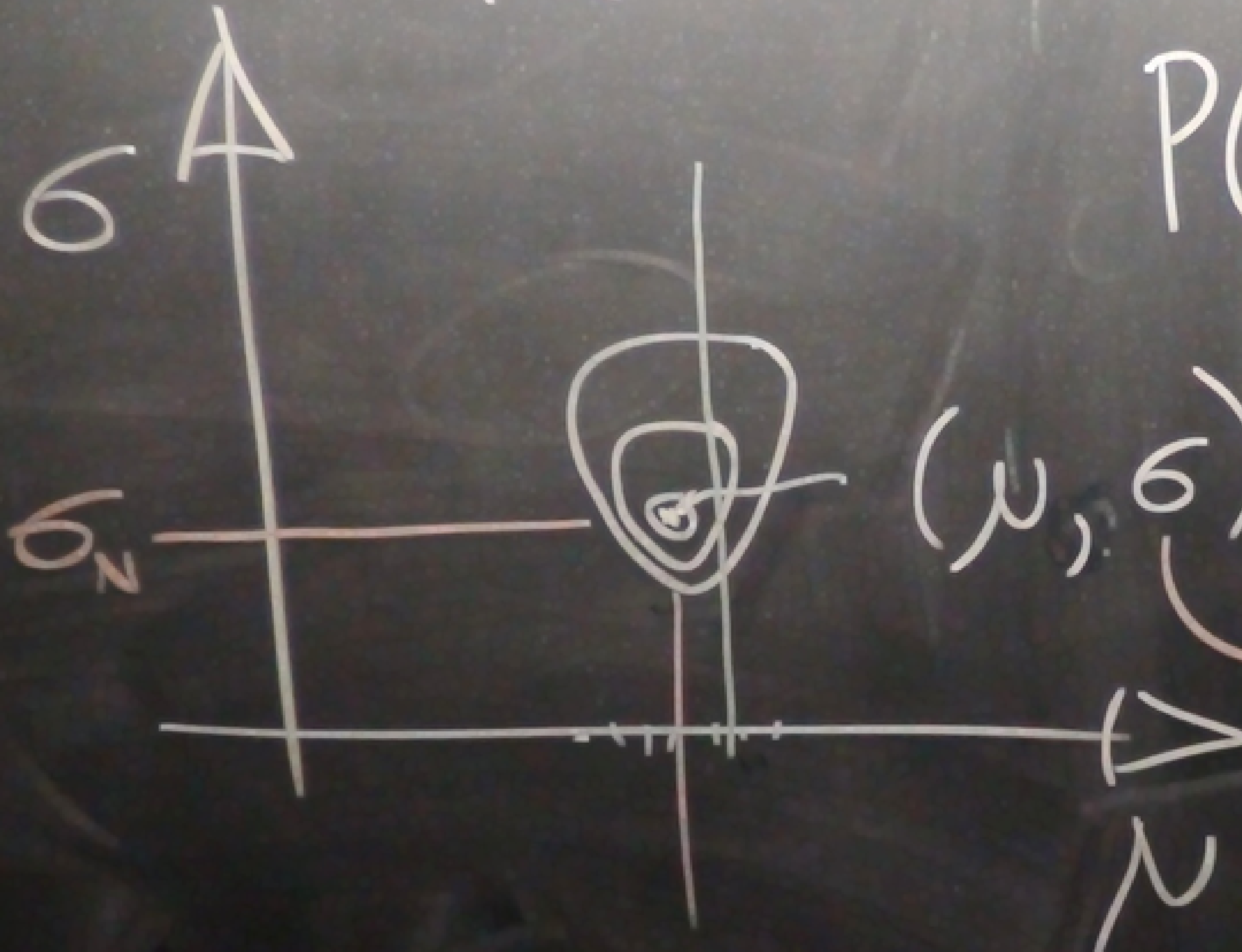

mu=1, sigma=0.445
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-6
-4
-2
0
2
4
6
0.05
0.04
0.03
0.02
0.01
0
-6
-4
-2
0
2
4
mu
0.01
0.1
1
10
sigma


0.01
0.001
0.0001
1e-05
1e-06
3
2
1
0
-1
-2
-3
mu
0.1
1
10
sigma


0.01
0.001
.0001
1e-05
1e-06
-3
-2
-1
0
1
2
3
mu
10
1
0.1
sigma

$N = 5$

$P(\{x_n\} \mid \mu, \sigma)$



$(\mu, \sigma)_{ML}$ → σ_N

$$\mu_{ML} = \bar{x} \equiv \frac{\sum_{n=1}^{N} x_n}{N}$$

Inference 1

$$P(\mu \mid \{x\}^N, \sigma) = \frac{P(\{x\} \mid \mu, \sigma)\, P(\mu}{P(\{x\} \mid \sigma)}$$

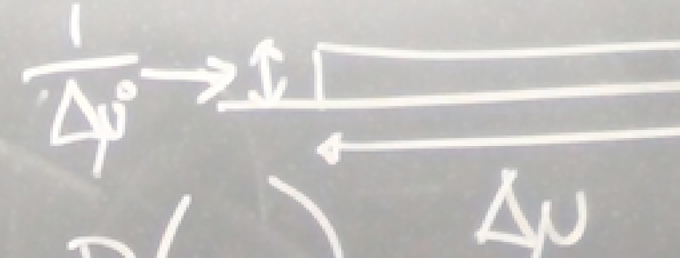
$\frac{1}{\Delta\mu} \rightarrow$

$\int d\mu \; P(\{x\} \mid \mu, \sigma)\, P($

Inference 2 $$P(\sigma \mid \{x\}, H) = \frac{P(\{x\} \mid \sigma)\, P(\sigma}{P(\{x\} \mid H)}$$

1

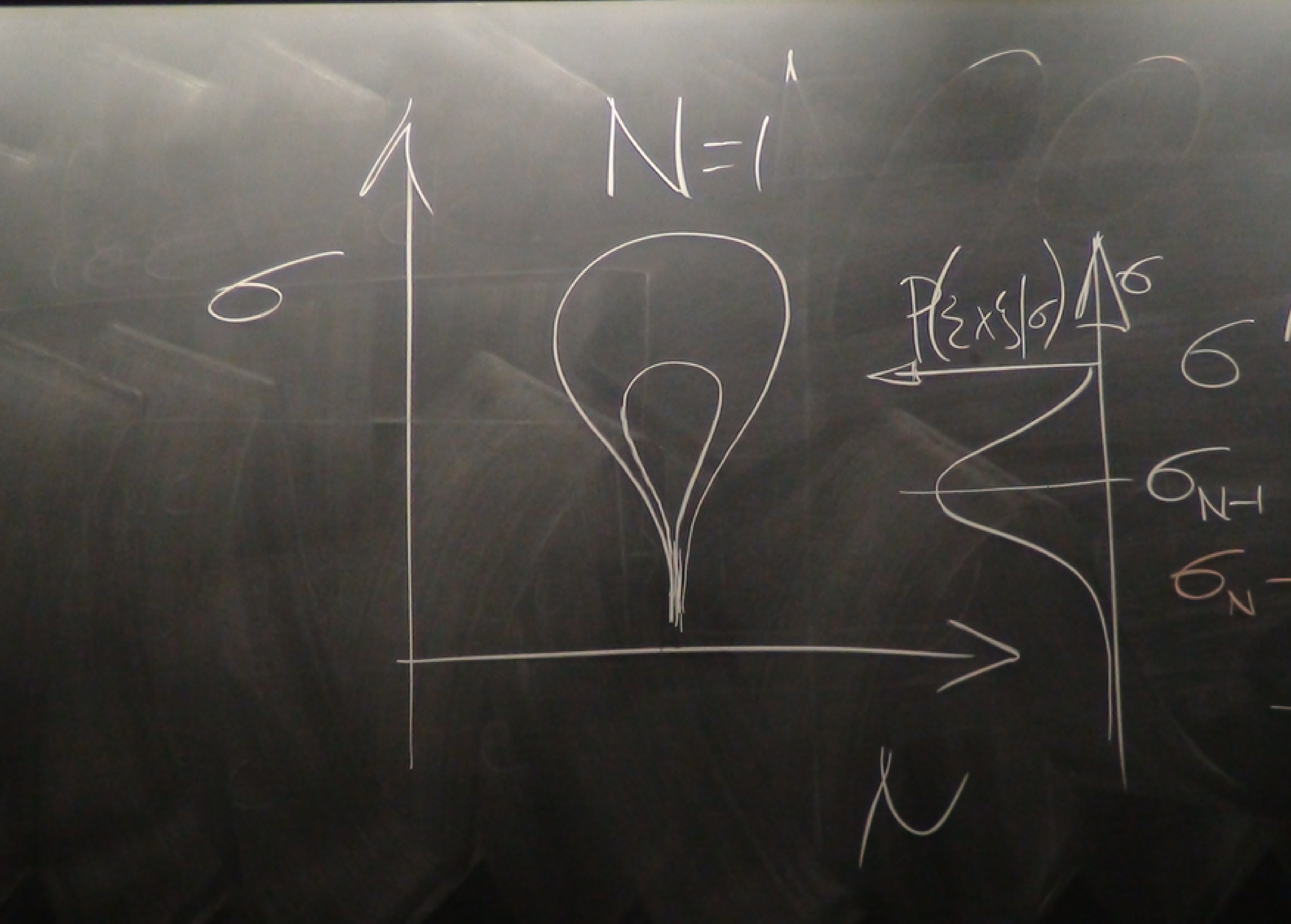
$$\mu \mid \{x\}_1^N, \sigma) = \frac{P(\{x\} \mid \mu, \sigma)\, P(\mu)}{P(\{x\} \mid \sigma)}$$



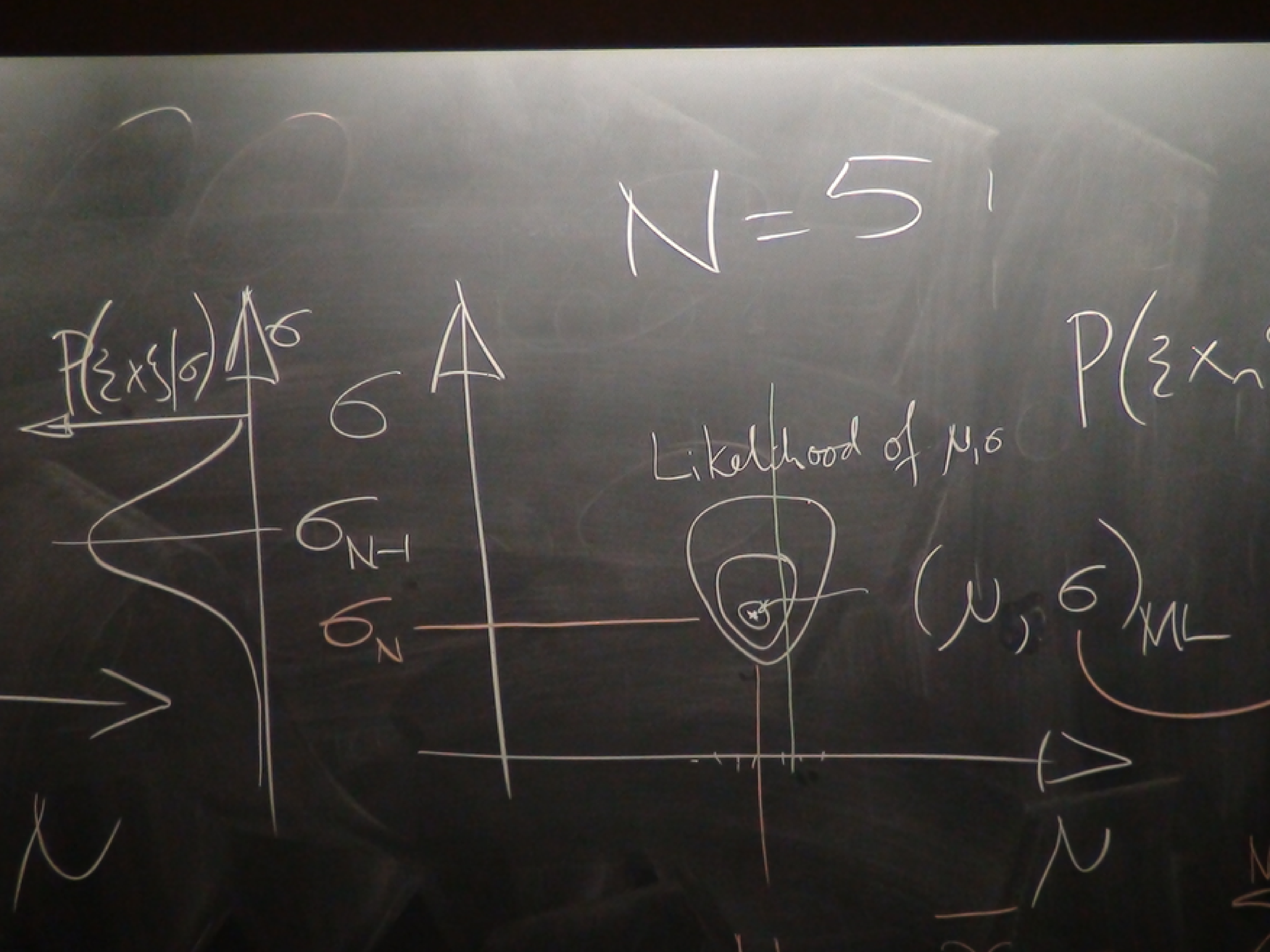
$$\int d\mu \; P(\{x\} \mid \mu, \sigma)\, P(\mu)$$

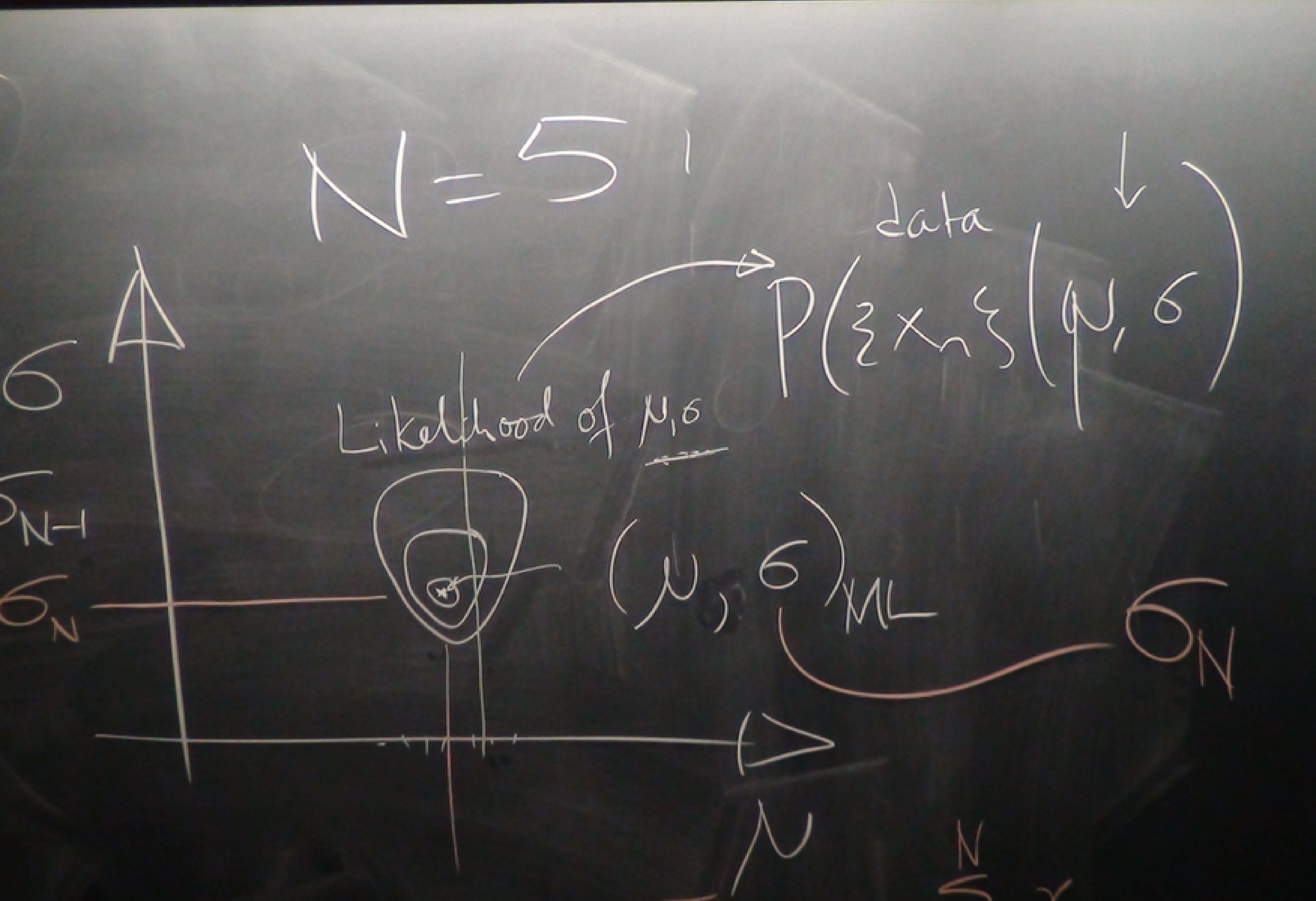
2

$$P(\sigma \mid \{x\}, H) = \frac{P(\{x\} \mid \sigma)\, P(\sigma)}{P(\{x\} \mid H)}$$

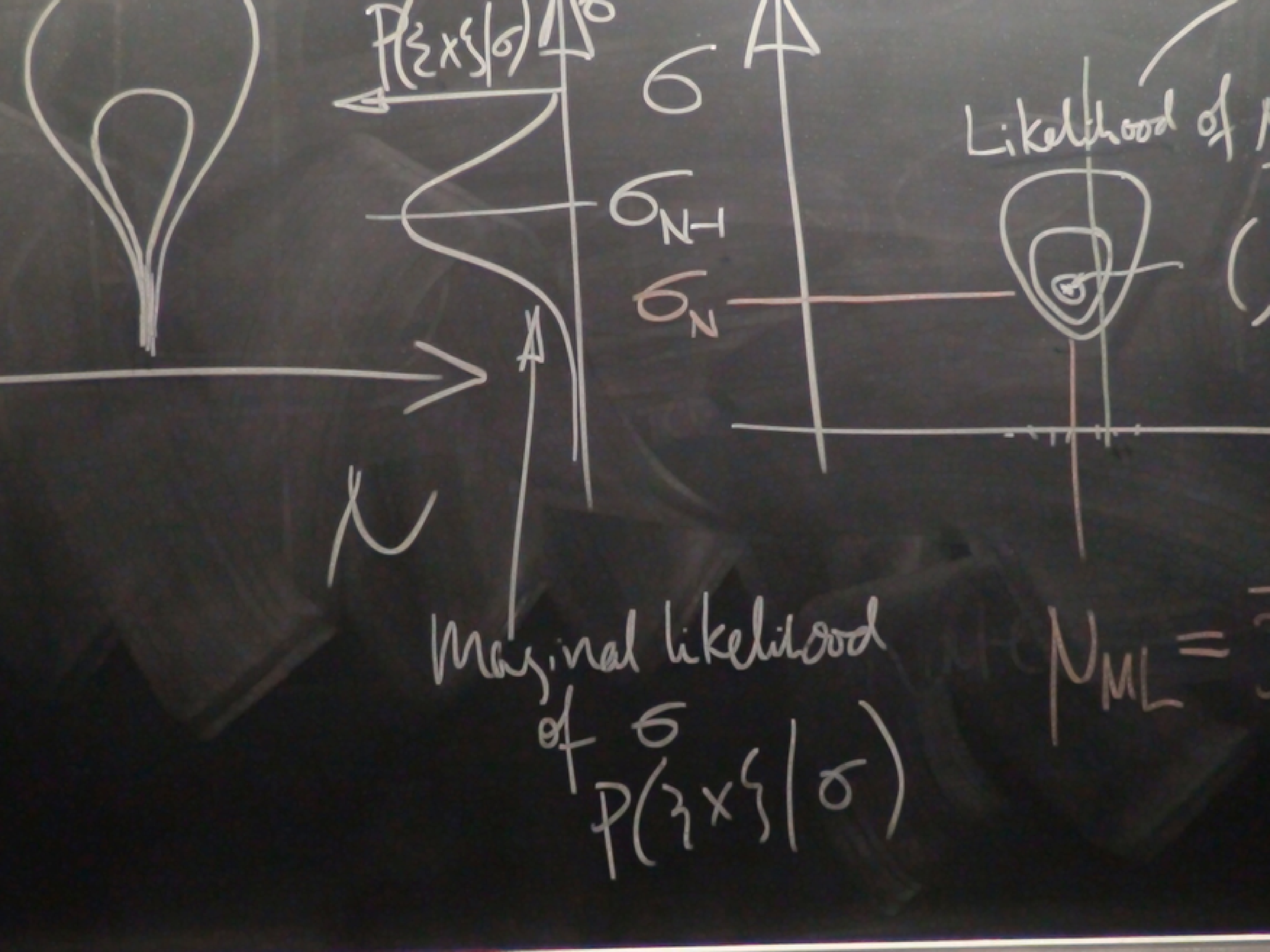


N=1
σ
P(ξ×ξ|σ)
σ
σ_{N+1}
σ_N

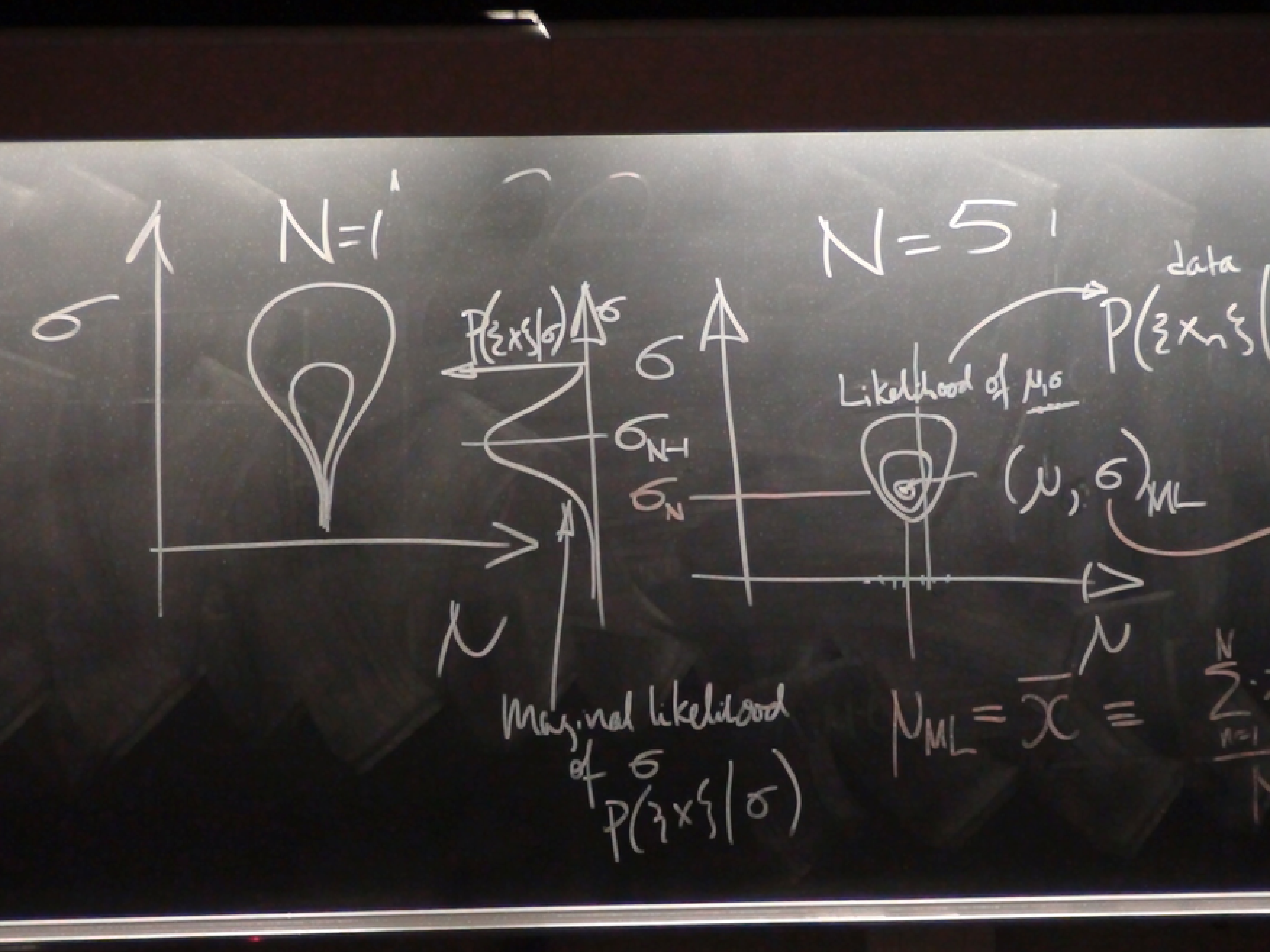

N = 5
P({x}|σ)
σ
σ
σ_{N-1}
σ_N
Likelihood of μ,σ
(μ,σ)_{ML}
P({x_n
μ
μ

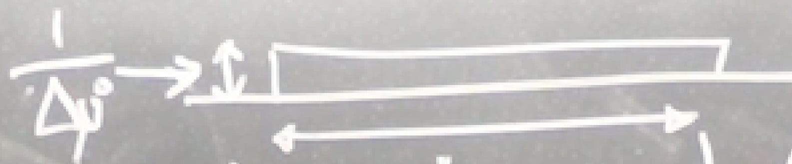


N = 5
data
P({x_n}|(μ, σ))
Likelihood of μ, σ
(μ, σ)_ML
σ_N
σ_{N-1}
σ
N



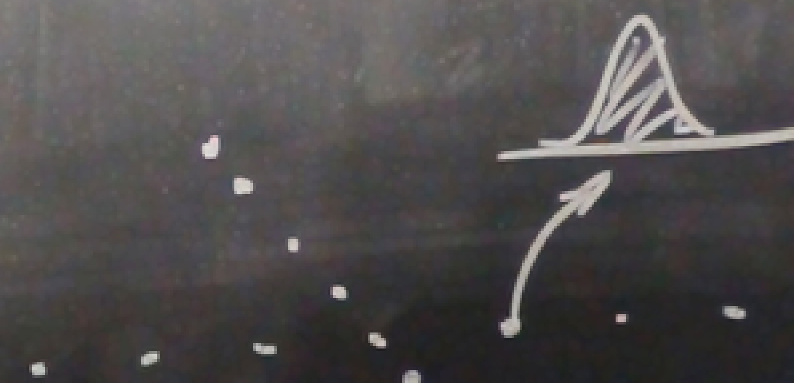
P({x}|σ)
σ
σ_{N-1}
σ_N
μ
Likelihood of μ
Marginal likelihood
of σ
P({x}|σ)
μ_{ML} =


N=1
σ
μ
P({x}|σ)
σ
σ_N-1
σ_N
Marginal likelihood
of σ
P({x}|σ)
N=5
data
P({x_n}|
σ
Likelihood of μ,σ
(μ, σ)_ML
μ
μ_ML = x̄ ≡



Infer μ, σ given

$$\frac{P(\{x\} \mid \mu, \sigma)\, P(\mu)}{P(\{x\} \mid \sigma)}$$



$$\int d\mu \; P(\{x\} \mid \mu, \sigma)\, P(\mu)$$

$$\ldots H) = \frac{P(\{x\} \mid \sigma)\, P(\sigma)}{P(\{x\} \mid H)}$$

Marginal likelihood of H
"Evidence for H"

Recommended homework

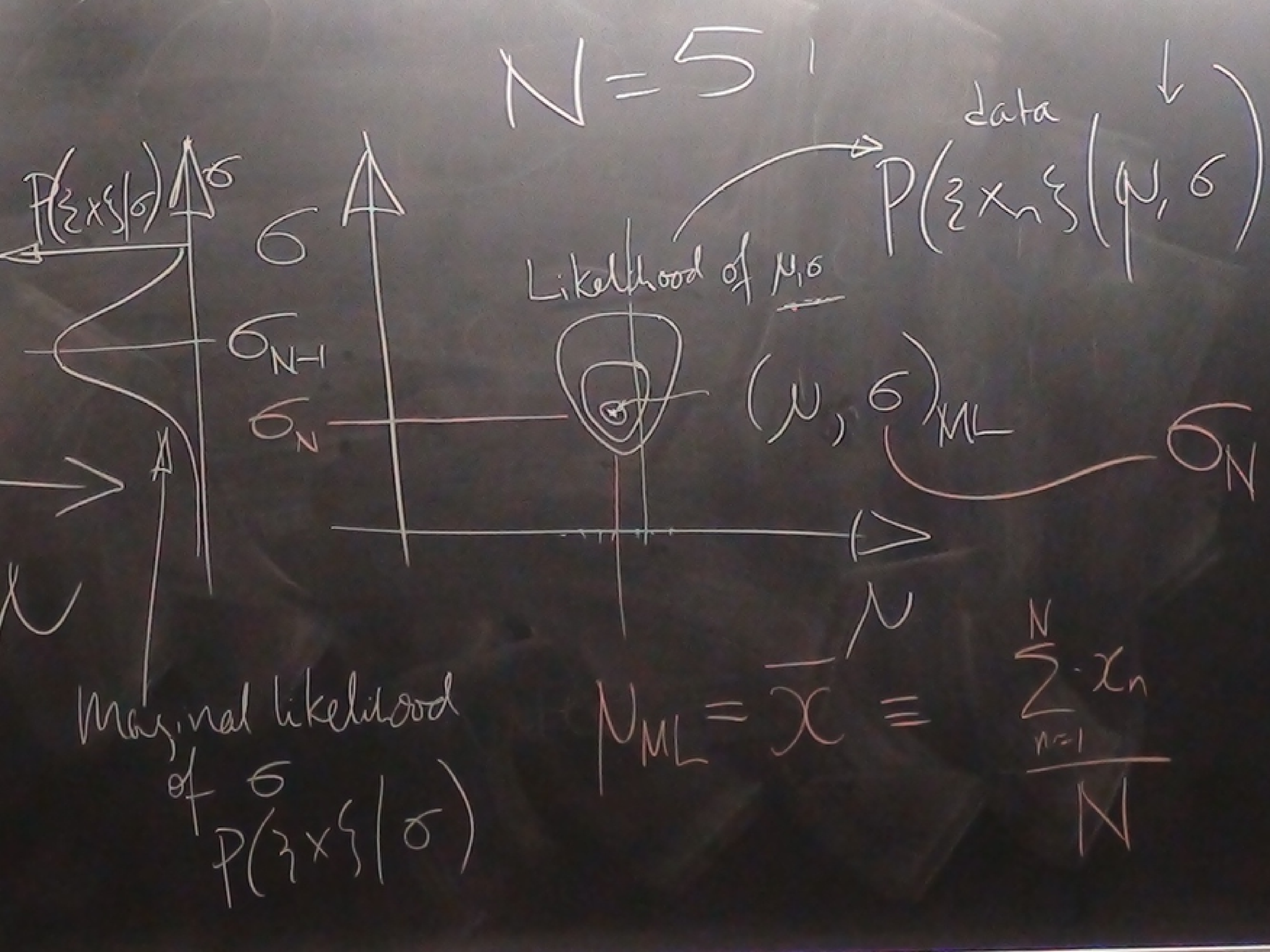
- Noisy channels - Chapters 8, 9, 10 (10.1-10.4 only)
  - Exercises 9.17 (p155); 10.12 (172); 15.12 (235)
  - and (if you want more practice) 15.11, 15.13, 15.15

- Invent a channel to pose to your colleagues:
  - 'what's the capacity of _this_?'

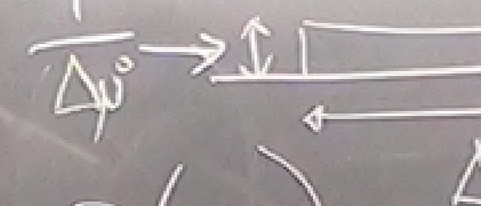
The reading associated with the current lectures is Chapters 3, 21 (especially sec 21.2), and 22 (especially sec 22.1), and 27.

Other recommended exercises are listed on handout 2.

www.inference.phy.cam.ac.uk/itprnn/


N = 5 !
data
P({x_n}|μ, σ)
P({x}|σ)
σ
σ_{N-1}
σ_N
Likelihood of μ, σ
(μ, σ)_ML
σ_N
μ
μ
Marginal likelihood
of σ
P({x}|σ)
μ_ML = x̄ ≡ Σ_{n=1}^{N} x_n / N

Inference 1



$$P(\mu \mid \{x\}_1^N, \sigma) = \frac{P(\{x\} \mid \mu, \sigma)\, P(\mu)}{P(\{x\} \mid \sigma)}$$

$$\int d\mu \; P(\{x\} \mid \mu, \sigma)\, P(\mu)$$

Inference 2

$$P(\sigma \mid \{x\}, H) = \frac{P(\{x\} \mid \sigma)\, P(\sigma)}{P(\{x\} \mid H)}$$