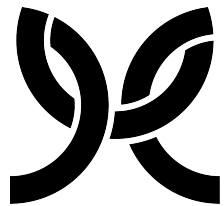


# **Chainsail: facilitating sampling of multimodal probability distributions**

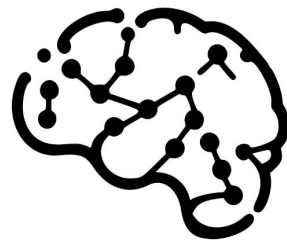
**Simeon Carstens**



**MODUS**



A Modus Create Company



# GENERATIVE AI

[www.tweag.io/group/genai](http://www.tweag.io/group/genai)

## MEMBERS



NOUR EL  
MAWASS



GUILLAUME  
DESFORGES



MARIA  
KNORPS



SIMEON  
CARSTENS



ALOÏS  
COCHARD

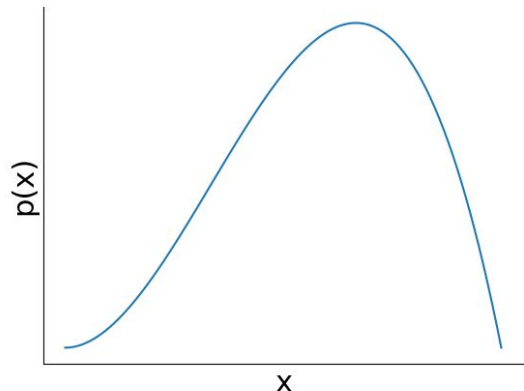


JOE NEEMAN

# Multimodal probability distributions



## Unimodal

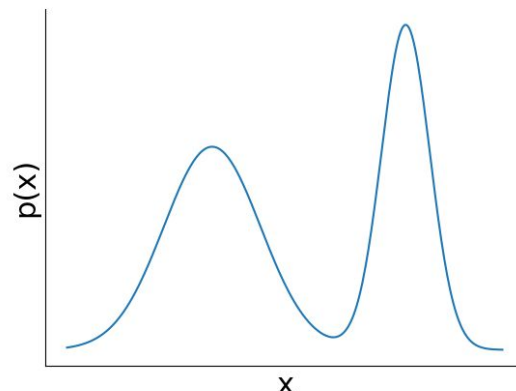


One single mode / bump / region with high probability

Examples:

- + Standard distributions (Gaussian, Beta, ...)

## Multimodal



Two or more modes / bumps / regions with high probability

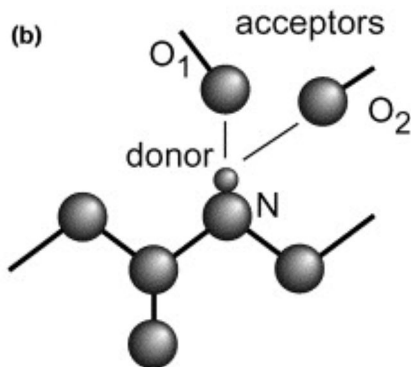
Examples:

- + Mixtures (e.g. Gaussian mixtures in soft k-means)
- + Bayesian inference:
  - + Ambiguous data
  - + Mismatch between prior distribution and likelihood

# Example: ambiguous data in protein structure



## Ambiguous distance restraints

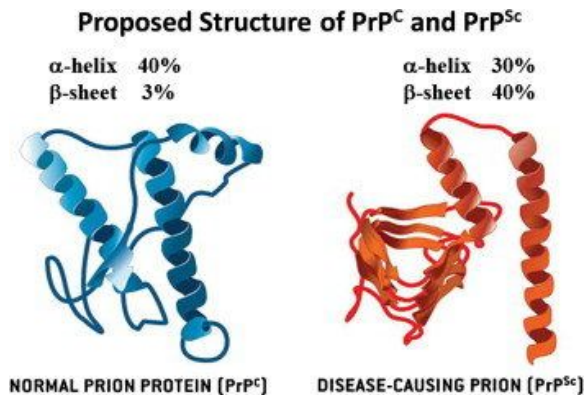


Data says that N is close to  $O_1$ ,  $O_2$  or both.  
Thus,

$p(x|D)$  is multimodal

x: 3D structure of protein, D: data

## Protein misfolding



Multiple folded states for the same protein

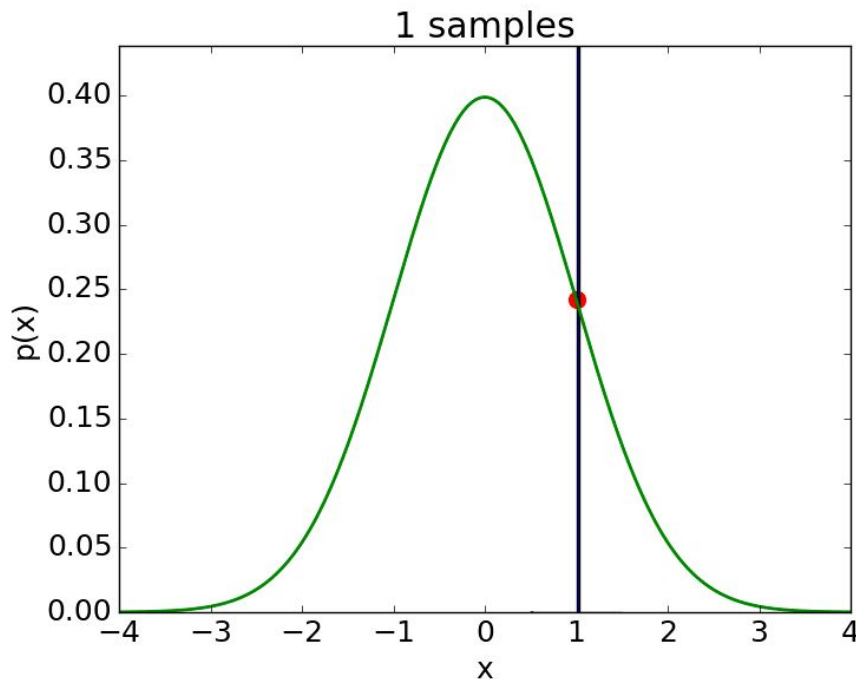
$p(x)$  is multimodal

“Wrong” folded state → neurodegenerative disease

# Sampling probability distributions: MCMC



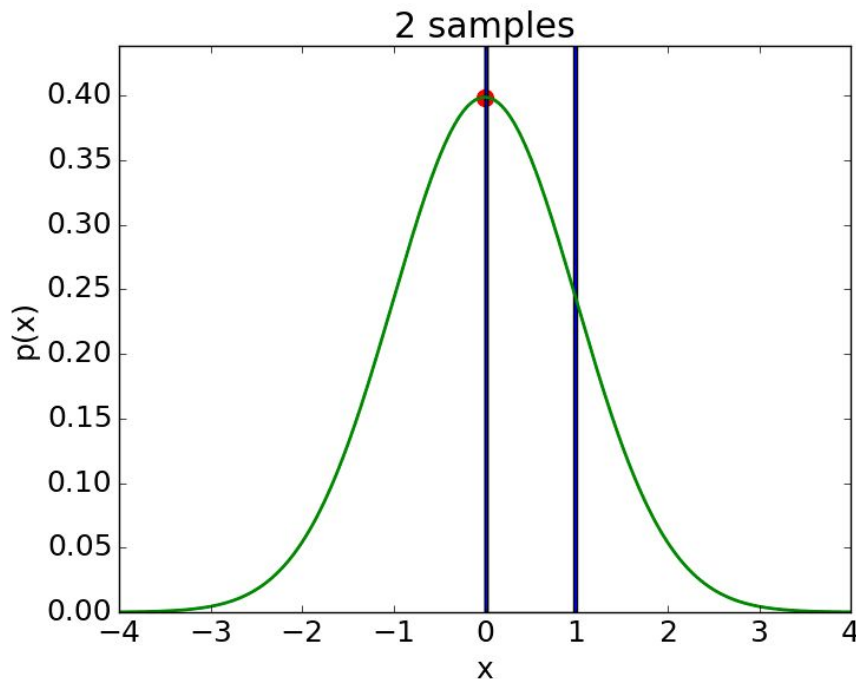
Sample arbitrary probability distributions with Markov chain Monte Carlo (MCMC):



# Sampling probability distributions: MCMC



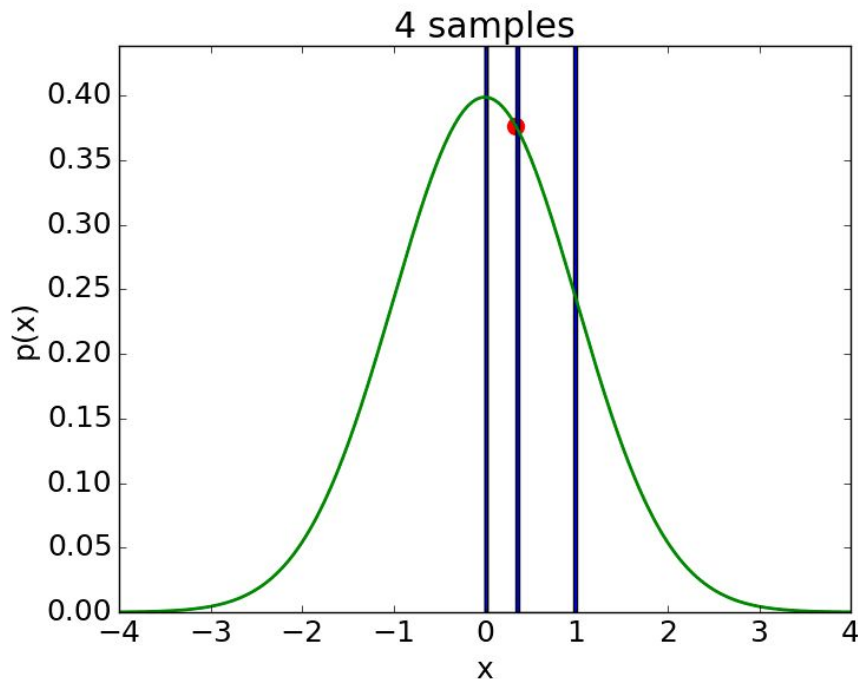
Sample arbitrary probability distributions with Markov chain Monte Carlo (MCMC):



# Sampling probability distributions: MCMC



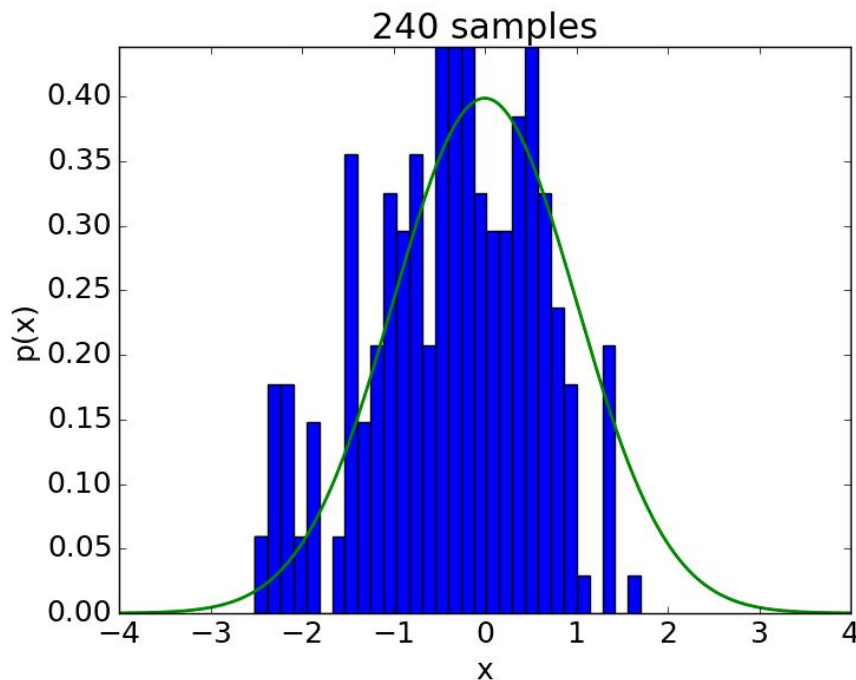
Sample arbitrary probability distributions with Markov chain Monte Carlo (MCMC):



# Sampling probability distributions: MCMC



Sample arbitrary probability distributions with Markov chain Monte Carlo (MCMC):

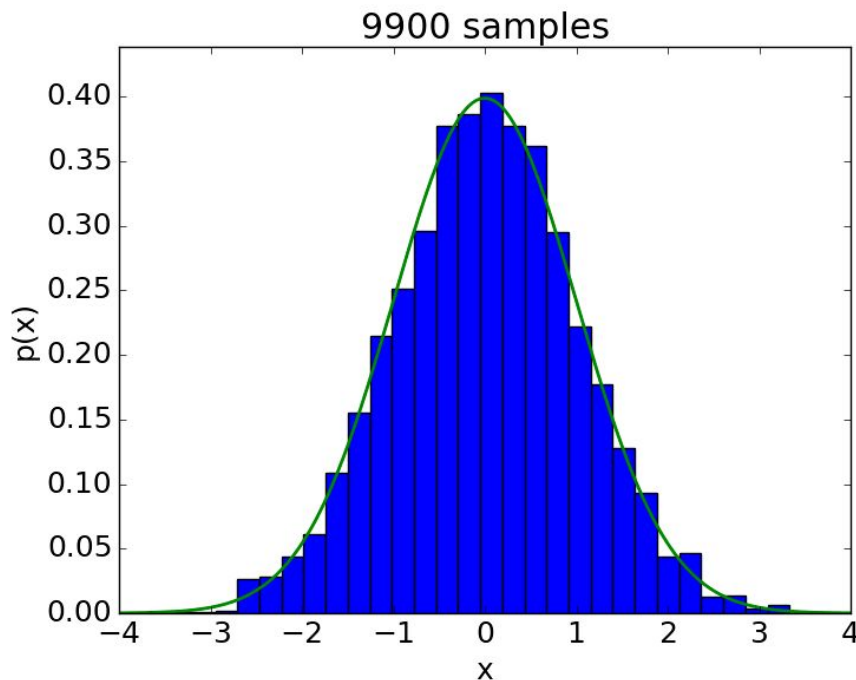




# Sampling probability distributions: MCMC



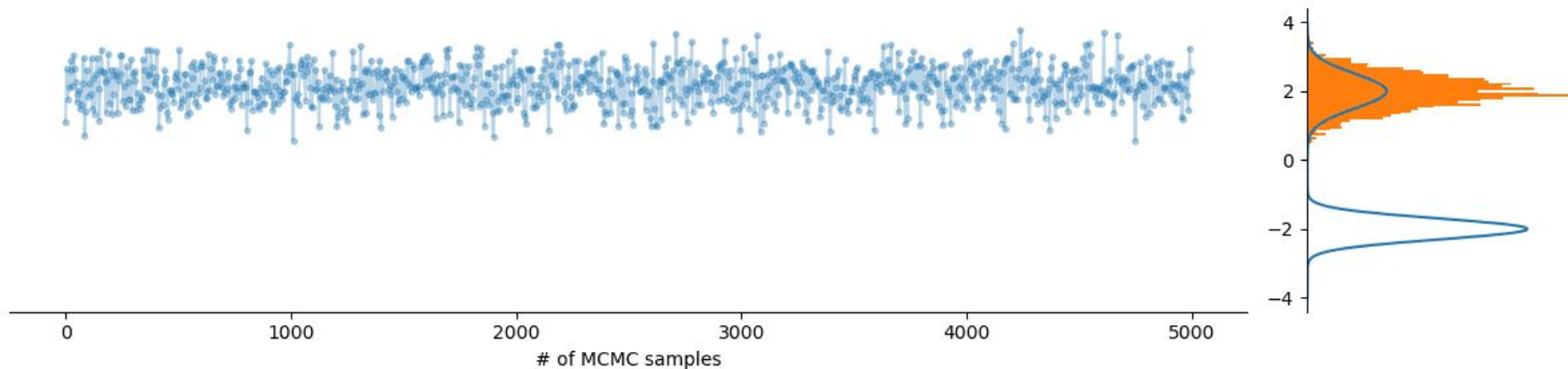
Sample arbitrary probability distributions with Markov chain Monte Carlo (MCMC):



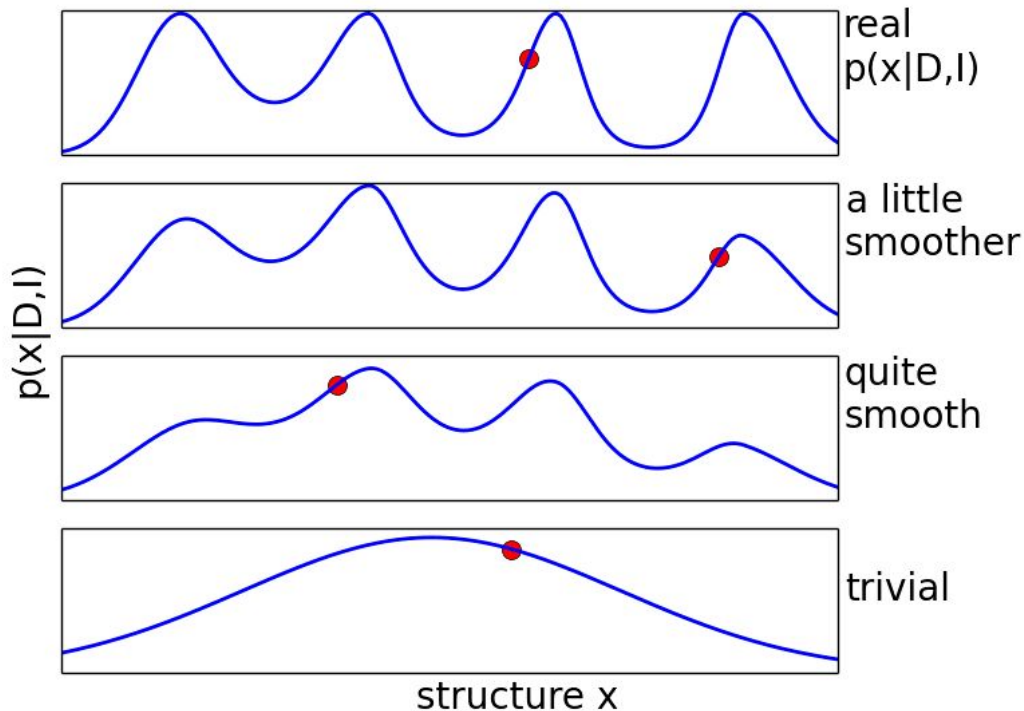
# Multimodality: a formidable foe



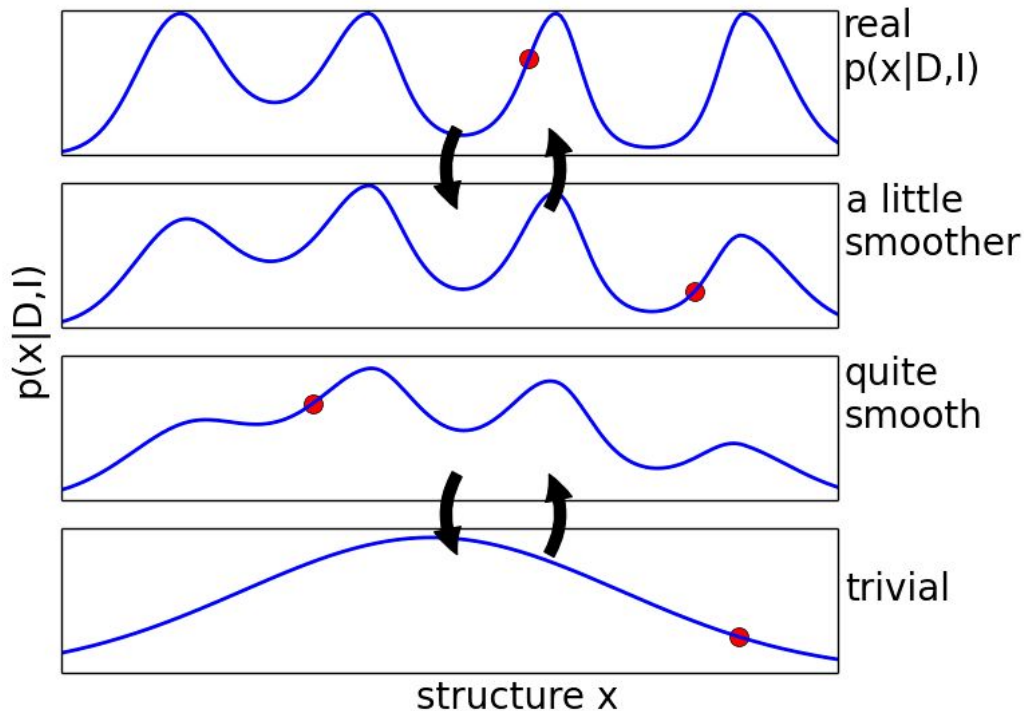
Chain can get trapped in a mode:



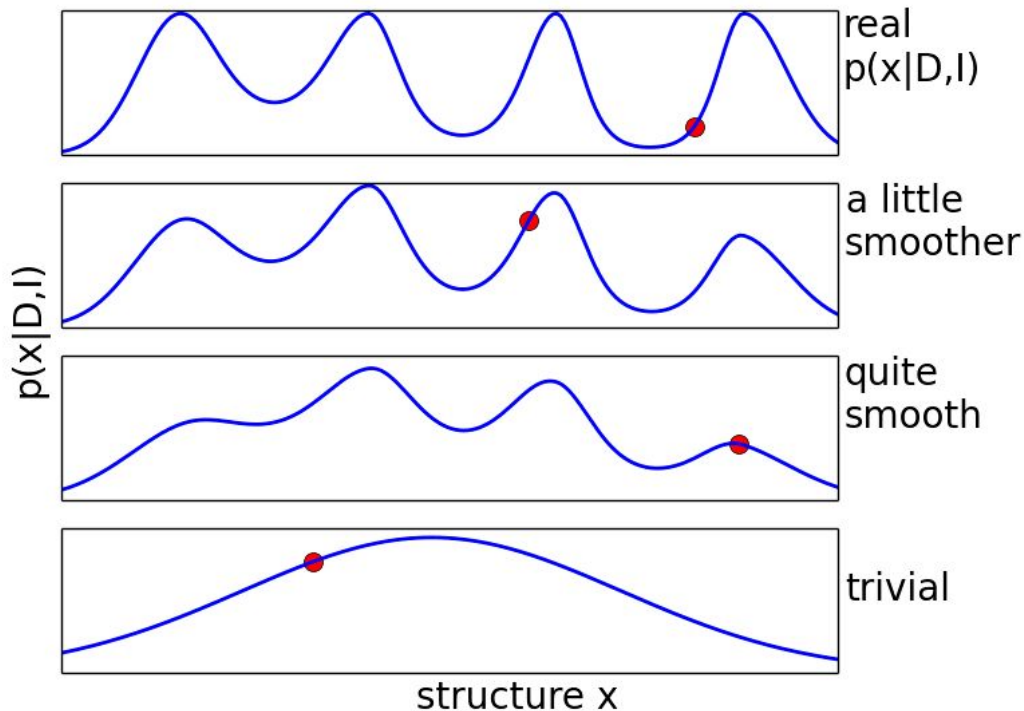
# Replica Exchange / Parallel Tempering



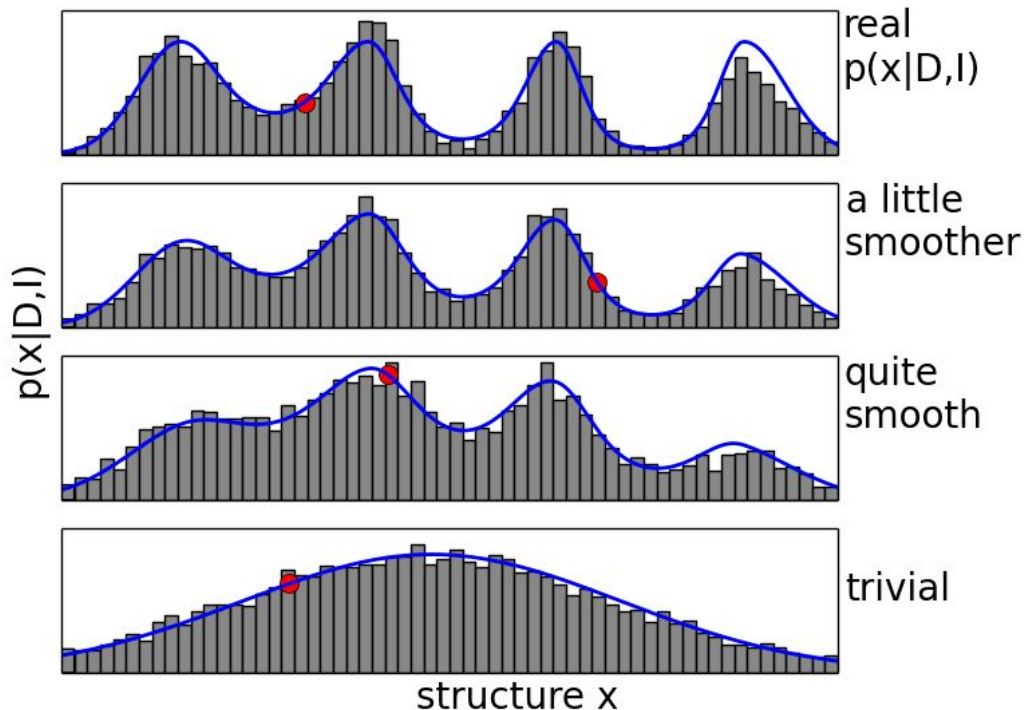
# Replica Exchange / Parallel Tempering



# Replica Exchange / Parallel Tempering



# Replica Exchange / Parallel Tempering

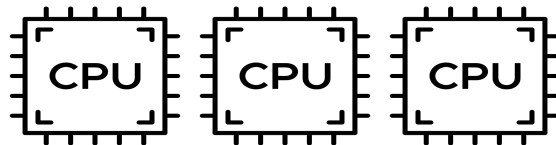


# Replica Exchange: problems



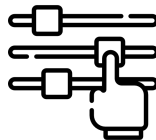
Multiple chains in parallel

→ parallel computing required



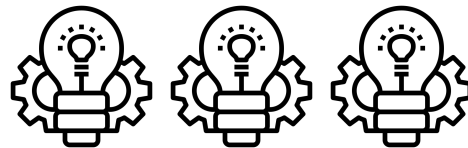
“Temperatures” difficult to tune

→ experience and trial and error required



Not implemented in probabilistic programming libraries

→ reinvent the wheel for your use case



# Chainsail: Replica Exchange on the



A Replica Exchange implementation that

- + interfaces with existing PPLs (PyMC, Stan)



- + Automatically finds good temperatures



- + Scales up and down automatically



- + Is easy to use

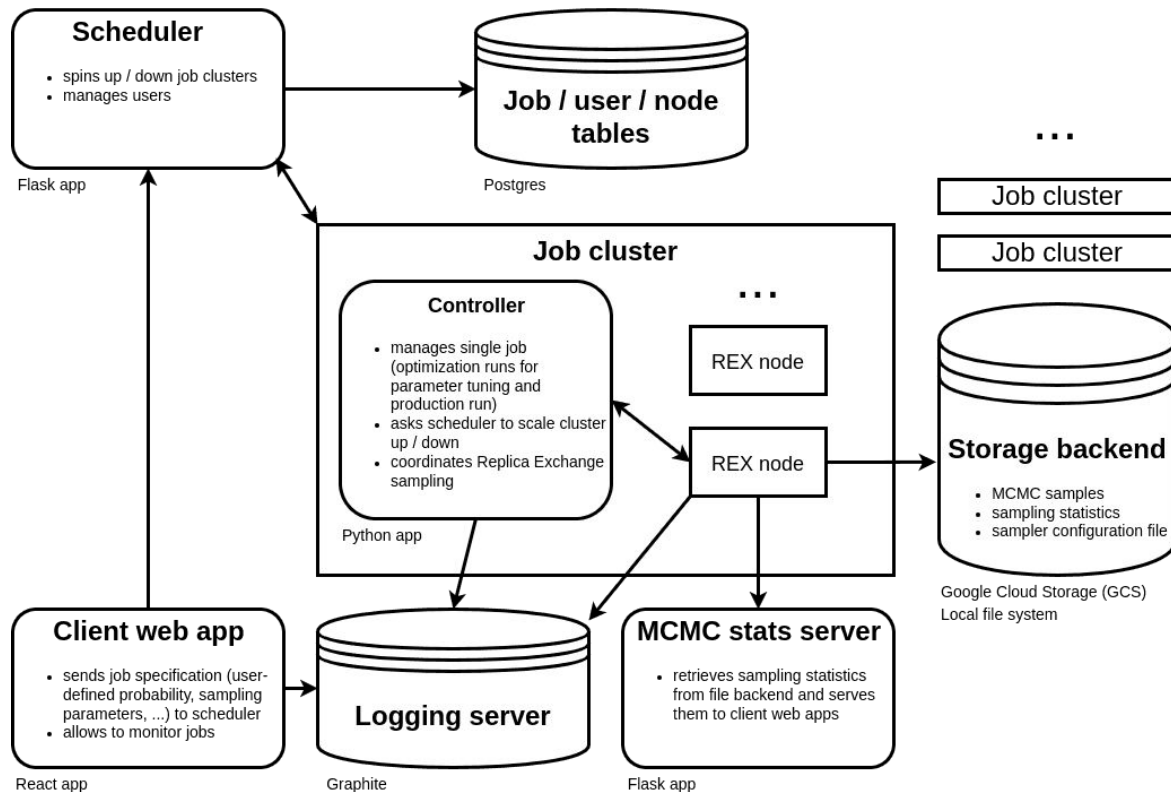


- + Runs barebone locally, full app locally, and in the cloud





# Chainsail: architecture



# Chainsail: tech stack



## Backend



PostgreSQL



OPEN MPI



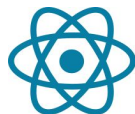
kubernetes



# Flask

web development,  
one drop at a time

## Frontend



# React

## NEXT.js

## Ops



Terraform



# Nix



minikube

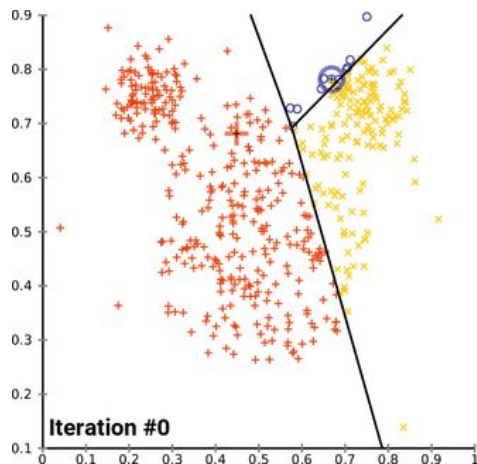


Google Cloud

# Example: soft k-means



## k-means



Iterate over:

1. Assign points to cluster with nearest mean
2. Recalculate means

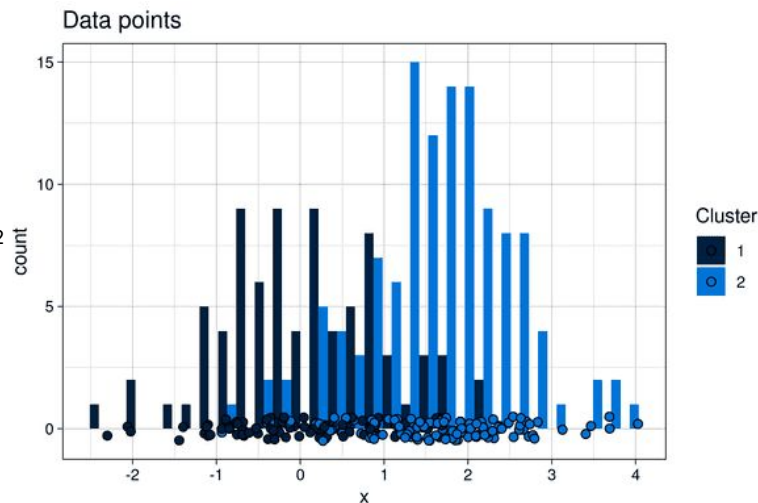
## Soft k-means

Probabilistic version of k-means: use Bayesian inference to infer cluster means

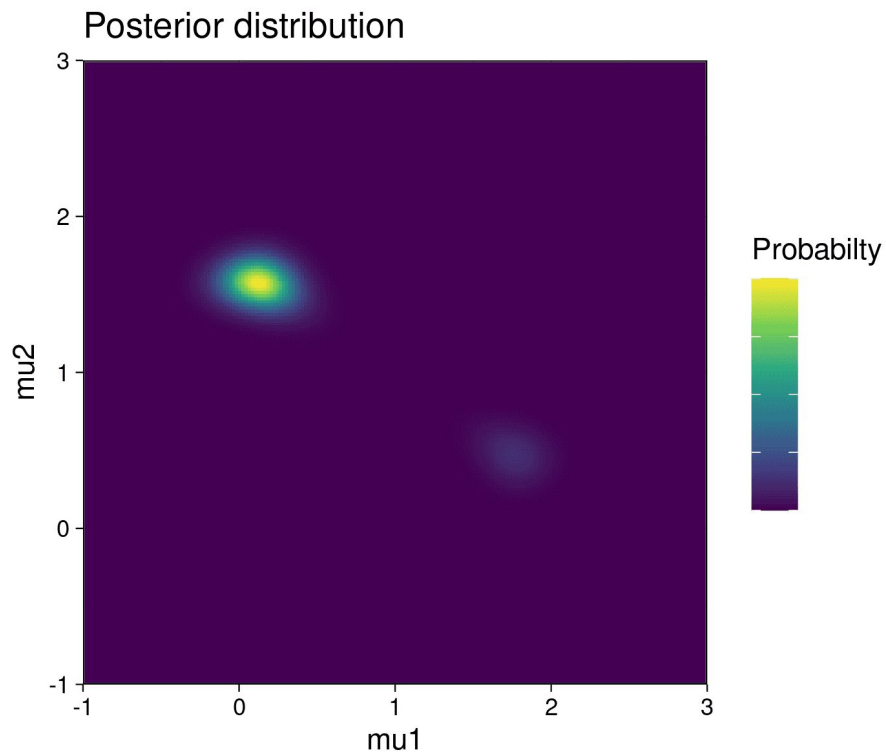
$$p(\mu_1, \mu_2 \mid \text{data}) \propto p(\text{data} \mid \mu_1, \mu_2) \times p(\mu_1, \mu_2)$$

Likelihood  $p(\text{data} \mid \mu_1, \mu_2)$ :

Assume that data was generated from Gaussian mixture with means  $\mu_1$  and  $\mu_2$  and **unequal weights**



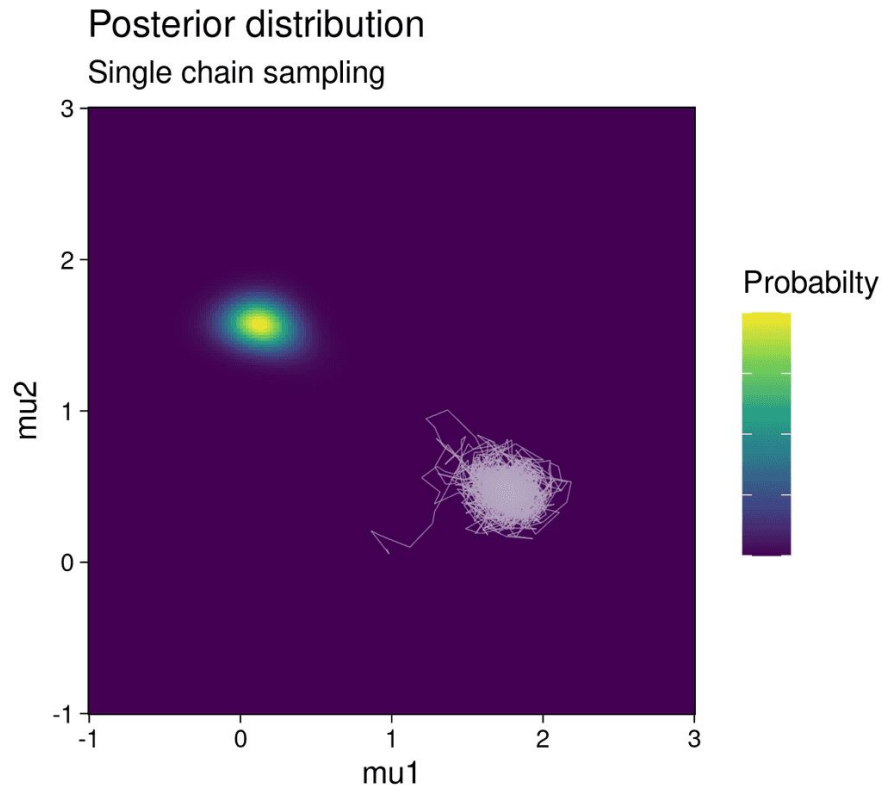
# Example: soft k-means



Bimodal posterior distribution for  
 $\mu_1, \mu_2$



# Example: soft k-means

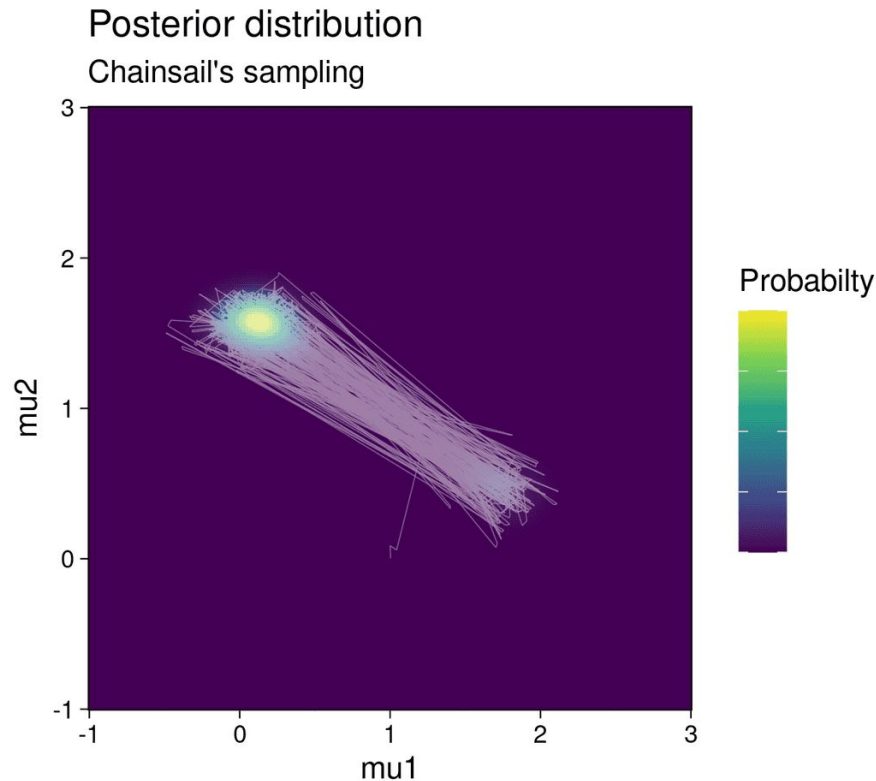


Bimodal posterior distribution for  $\mu_1, \mu_2$

Hard to sample with one MCMC chain, but...



# Example: soft k-means



Bimodal posterior distribution for  $\mu_1, \mu_2$

Hard to sample with one MCMC chain, but...

... a piece of cake with Chainsail!



# Example: soft k-means

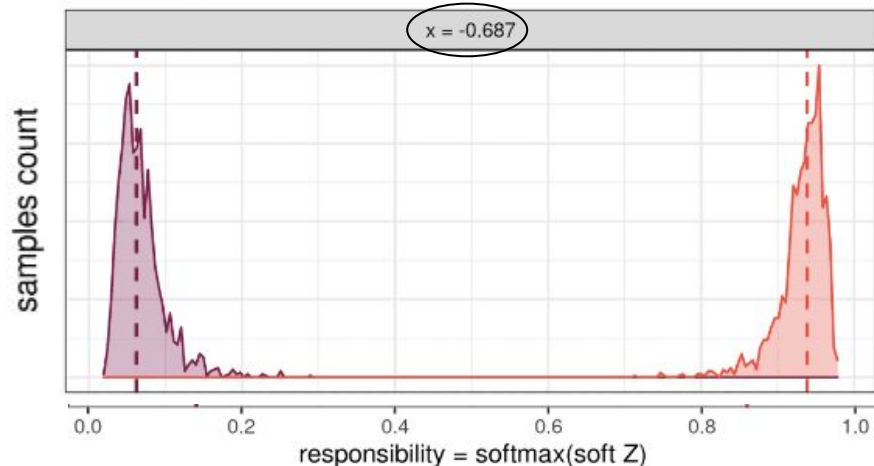


Reminder:

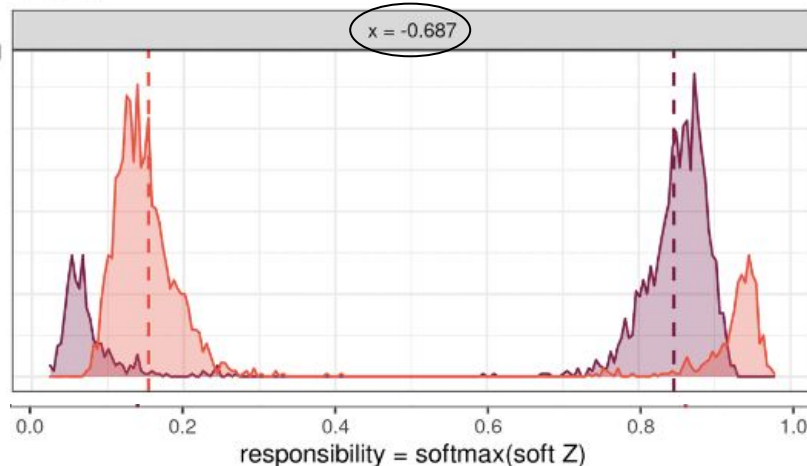
- + cluster 1 is centered at 0
- + cluster 2 is centered at 1.5

**Responsibility:** measure of how strongly a cluster explains a data point

Single chain



Chainsail



Wrong cluster assignments:

- Low responsibility for cluster 1
- High responsibility for cluster 2

Correct cluster assignments:

- High responsibility for cluster 1
- Low responsibility for cluster 2
- ... and a small probability for the inverse!

# Current issues / things to do



## Big picture:

- + Well, they didn't 🐱



**Real-world use cases needed - reach out to us!**

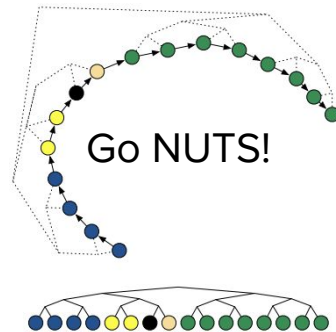
- + Little resources to continue development, but happy to maintain!

## Features / functionality:

- + Only very basic single-chain sampling implemented (naïve Hamiltonian Monte Carlo)
- + Temperature auto-tuning works well only for “simple” cases

## Code / technical debt:

- + Code is rotting (Docker images need constant updating, Python dependencies break, ...)





# Credits



The Chainsail team:

Dorran Howell

Saeed Hadikhanloo

Etienne Jean

... and many others 🧡

**Thank you**  
for your attention!

Attributions:

- + Icons: [flaticon.com](https://flaticon.com)
- + NUTS figure:  
“The No-U-Turn Sampler [...]”,  
Hoffman, M., and Gelman, A.,  
[arXiv](https://arxiv.org/abs/2011.01789), 2011
- + Ambiguous distance restraints  
figure:  
Nilges, M: “Ambiguous distance  
data [...]”, *Folding and Design*,  
1997
- + Prion figure:  
“Prion Diseases as  
Transmissible Zoonotic  
Diseases”, Lee, J. et al., *Osong  
Public Health and Research  
Perspectives*, 2013
- + Rotten fruit: [istockphoto.com](https://istockphoto.com)
- + “Build it...” meme: [imgflip.com](https://imgflip.com)

# Resources



[github.com/tweag/chainsail](https://github.com/tweag/chainsail)

(code)

[github.com/tweag/chainsail-resources](https://github.com/tweag/chainsail-resources)

(examples, documentation)



[tweag.io/blog](https://tweag.io/blog) -> 🔍 “chainsail”

<http://chainsail.io>

(mostly links...)