Von Neumann's Roulette: Spinning the Wheel with MCMC

Ricardo Chávez Murillo

DAWGI, May 2024

Monte Carlo Methods

Ulam & von Neumann (1940s)

- Developed by Stanislaw Ulam and John von Neumann in the 1940s.
- Inspired by random processes observed in games of chance, named after the Monte Carlo Casino.
- First practical application during the Manhattan Project for simulating neutron diffusion.
- Method involves using random sampling to solve problems estimating probabilities, calculating integrals.
- Significance: Transforms complex, deterministic problems into simpler, probabilistic simulations.

MCMC Methods

The Metropolis et al. (1953) paper

- Fundamental Markov Chain Monte Carlo (MCMC) method for sampling from complex probability distributions.
- Proposal Distribution: Generates a candidate state by modifying the current state.
- Acceptance Criterion: Accepts the candidate state based on a probability ratio.
- Rejection: If rejected, retains the current state and counts it again in the sampling.
- Repeatedly applies these steps to explore the state space, gathering samples.
- Key Point: Enables efficient exploration of large, complex state spaces by sampling more frequently from regions with higher probabilities.

$$\mathfrak{I} = \int F(\theta) \exp\{-E(\theta)/kT\}\, d\theta \Big/ \int \exp\{-E(\theta)/kT\}\, d\theta,$$

$$E(\theta) = \frac{1}{2} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} V(d_{ij})$$

$$Z(T) = \int \exp\{-E(\theta)/kT\}\, d\theta$$

$$x_i' = x_i + \sigma \xi_{1i} \quad \text{and} \quad y_i' = y_i + \sigma \xi_{2i}$$

$$\min\{1, \exp(-\Delta E/kT)\}$$

MCMC Methods

The Hastings (1970) paper

- Enables sampling from a broader range of distributions by modifying the proposal distribution.

- **Asymmetric Proposal Distributions**: Introduces a proposal function q(x|y) for moving from state x to y and q(y|x) for the reverse, allowing for non-symmetrical proposals.

- **Acceptance Probability**: Adjusted to min(1, \frac{p(y)q(x|y)}{p(x)q(y|x)}) to maintain detailed balance and ensure the Markov chain converges to the target distribution.

- Widely used in statistics, physics, and computational biology for complex probabilistic models.

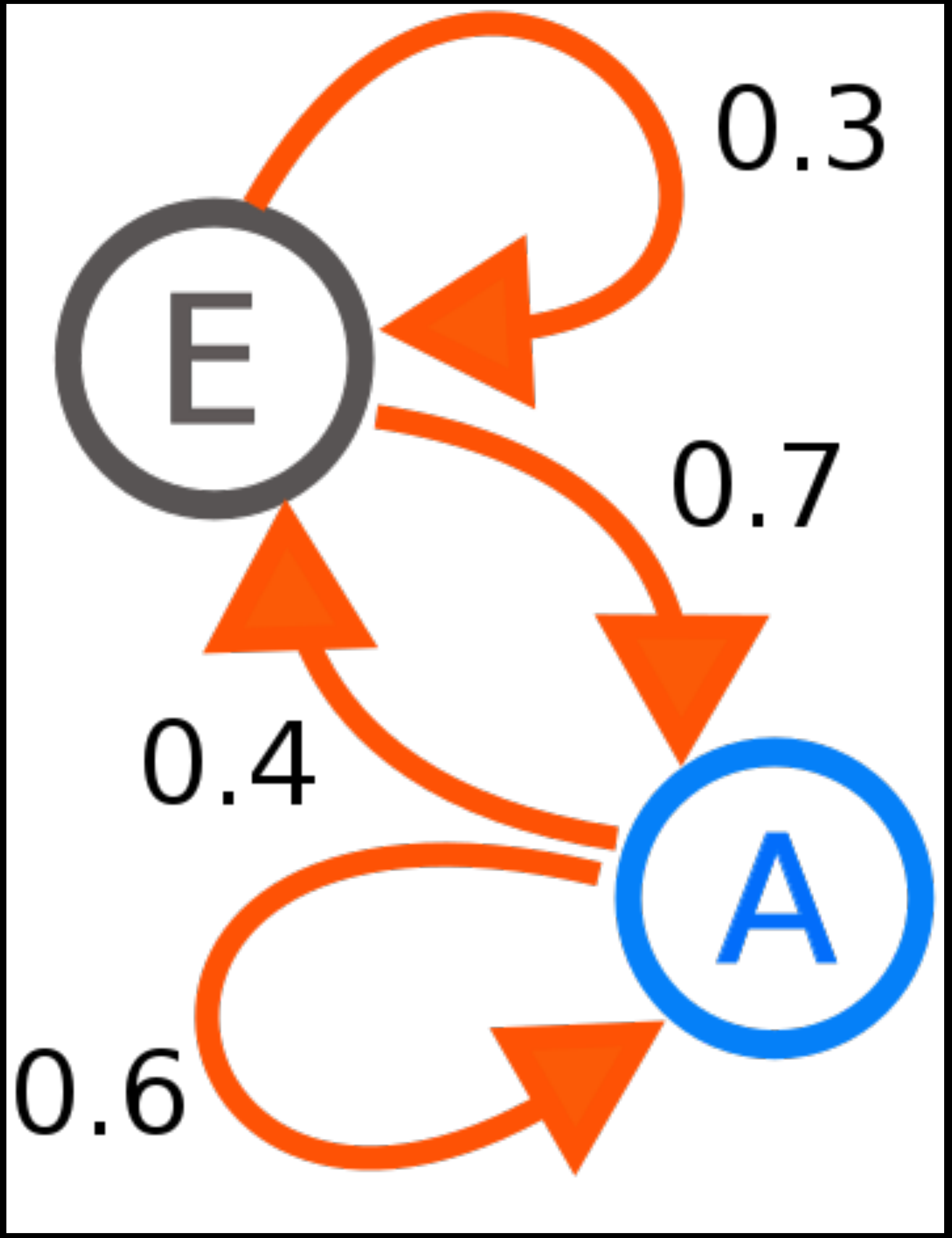
- **Key Takeaway:** Expands the applicability of MCMC methods by accommodating more complex, realistic proposal distributions.

$$\alpha_{ij} = \frac{s_{ij}}{1 + (\pi_i/\pi_j)(q_{ij}/q_{ji})}$$

MCMC Methods

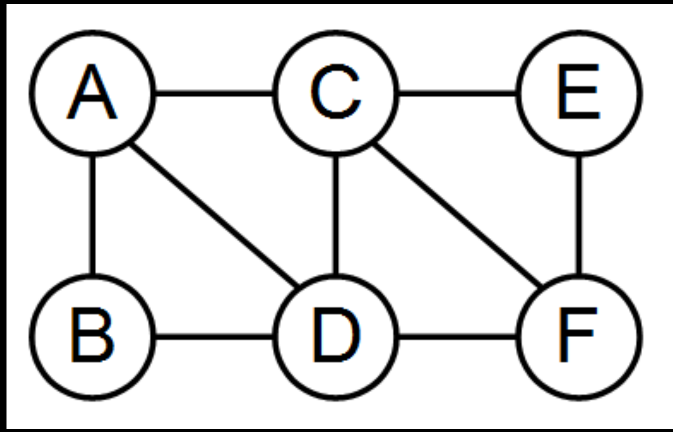
Exploring Markov Chains: A Fundamental Concept in Stochastic Processes

- A Markov Chain is a stochastic process that transitions from one state to another within a set of possible states.
- **Markov Property:** The future state depends only on the current state, not on the path taken to reach it (memorylessness).
- **States:** Each state represents a possible condition or status of the system at a given time.
- **Transition Probability:** Defined for each pair of states, these probabilities dictate the likelihood of moving from one state to another.
- **Applications:** Used in a variety of fields including finance, genetics, computer science, and more, to model systems where outcomes are random but contextually dependent.
- **Key Takeaway:** Markov Chains provide a powerful framework for modelling random systems where the future is conditionally dependent only on the present state.

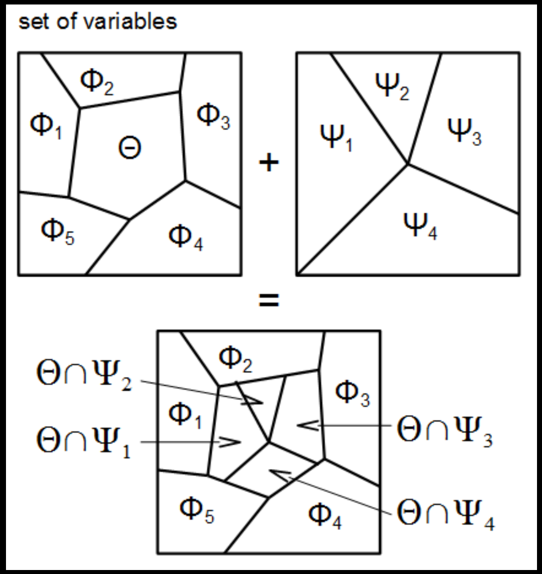


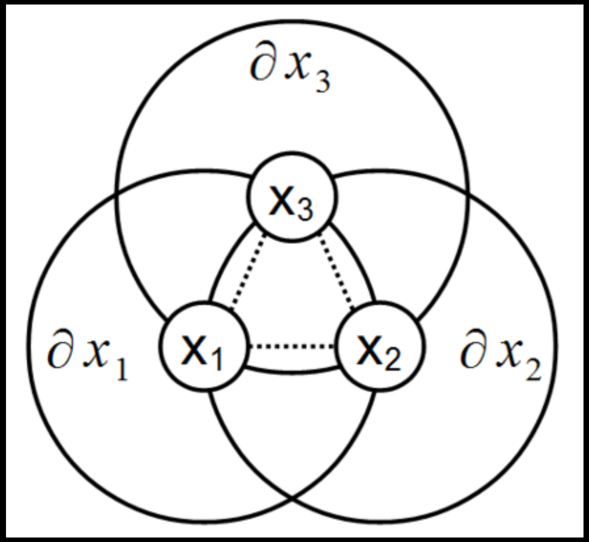
MCMC Methods

The Hammersley-Clifford theorem



- Fundamental theorem in the field of Markov random fields and graphical models.
- Established by J.M. Hammersley and P. Clifford in the early 1970s.
- **Core Principle**: A probability distribution adheres to a Markov property relative to an undirected graph if and only if it can be represented as a Gibbs distribution over the graph.
- **Implications:** The theorem provides a crucial link between graph theory and probability distributions, stipulating that the independence of random variables within a graph is dictated by the absence of edges connecting them.
- Widely applied in statistical physics, computer vision, and Bayesian networks to enforce local dependencies in complex systems.
- **Key Takeaway:** The theorem's utility lies in simplifying complex dependency structures in probabilistic models, allowing for more tractable analyses and applications.

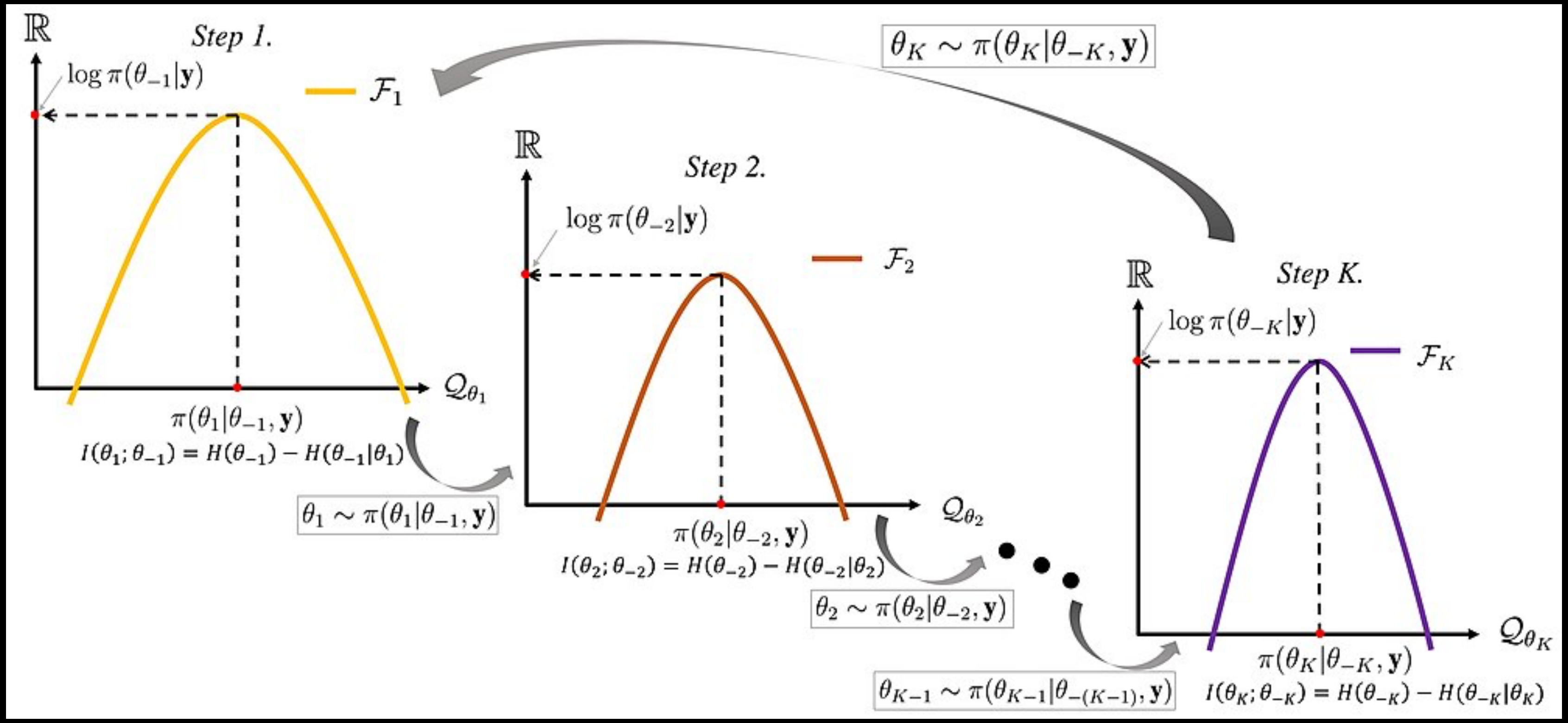




MCMC Methods

Gibbs Sampling

- Gibbs Sampling is a specialised case of the Metropolis-Hastings algorithm used extensively in Bayesian statistics.
- Operates by sequentially sampling from the conditional distributions of each variable while holding all other variables fixed.
- **Process Overview**: Iteratively updates each variable in turn, using the latest values of other variables to compute the conditional distributions.
- **Advantages**: Simplifies computation by avoiding the need to compute complex acceptance ratios as in general MCMC methods.
- Effectively handles high-dimensional data and complex models, particularly useful in hierarchical Bayesian models.
- **Key Takeaway:** Gibbs Sampling is pivotal for efficiently exploring high-dimensional probability distributions, facilitating more practical and scalable Bayesian inference.



MCMC Methods

Central Limit Theorem in Markov Chain Monte Carlo

- The CLT states that the mean of a sufficiently large number of independent random variables, each with a well-defined expected value and variance, will be approximately normally distributed.

- In MCMC, the CLT assures that under certain conditions, the sample averages of a function of the Markov chain states converge to a normal distribution as the number of samples increases.

- **Conditions for CLT in MCMC:** The Markov chain must be ergodic, ensuring it converges to its stationary distribution regardless of the initial state.

- **Implications:** Allows for quantification of the sampling error in MCMC simulations, enabling the use of confidence intervals and hypothesis testing.

- Critical for the practical application of MCMC methods in statistical inference, ensuring the reliability of estimates derived from Markov chains.

- **Key Takeaway:** The Central Limit Theorem provides foundational assurance for using MCMC methods in complex Bayesian analyses, highlighting the robustness of these techniques in statistical estimation.

$$\frac{\hat{f}_n^{\text{MCMC}} - \mathbb{E}_p[f(\theta)]}{\text{MCMC-SE}_n[f]} \sim \mathcal{N}(0,1)$$

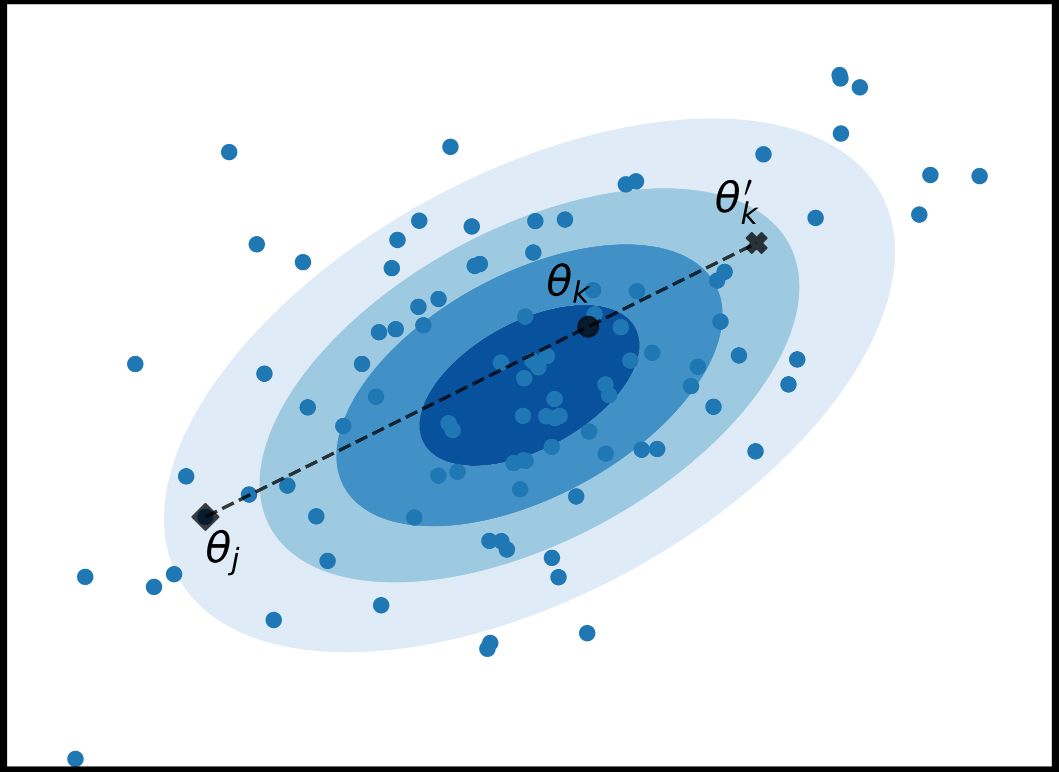
$$\text{MCMC-SE}_n[f] = \sqrt{\frac{\text{Var}_p[f]}{\text{ESS}_n[f]}}$$

$$\text{ESS}_n[f] = \frac{n}{\tau[f]}$$

MCMC Methods

Ensemble MCMC Methods

- Ensemble MCMC methods utilise multiple, interacting Markov chains to sample from a target distribution, improving exploration and convergence.
- **Parallel Chains:** Operates several chains in parallel, each exploring different regions of the parameter space, which can share information to avoid local minima.
- **Diversity in Sampling:** Ensures diverse sampling by maintaining a population of solutions, reducing the risk of convergence to local optima.
- **Examples:** Popular examples include the Affine Invariant Ensemble Sampler (used in emcee) and the Sequential Monte Carlo Samplers.
- **Applications:** Widely used in fields requiring robust exploration of complex probability landscapes, such as astrophysics, statistical physics, and Bayesian model fitting.
- **Key Takeaway:** Ensemble MCMC methods significantly increase the efficiency and reliability of MCMC algorithms, especially in complex, multi-modal problems.

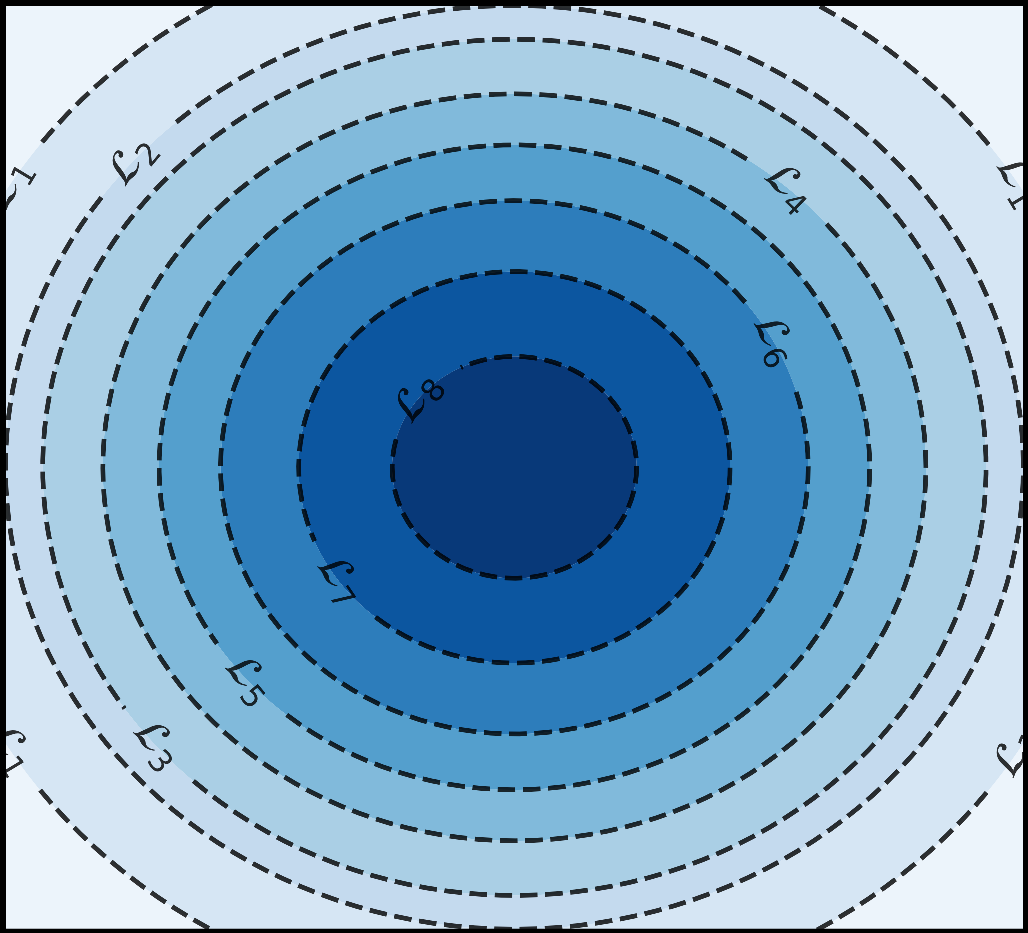


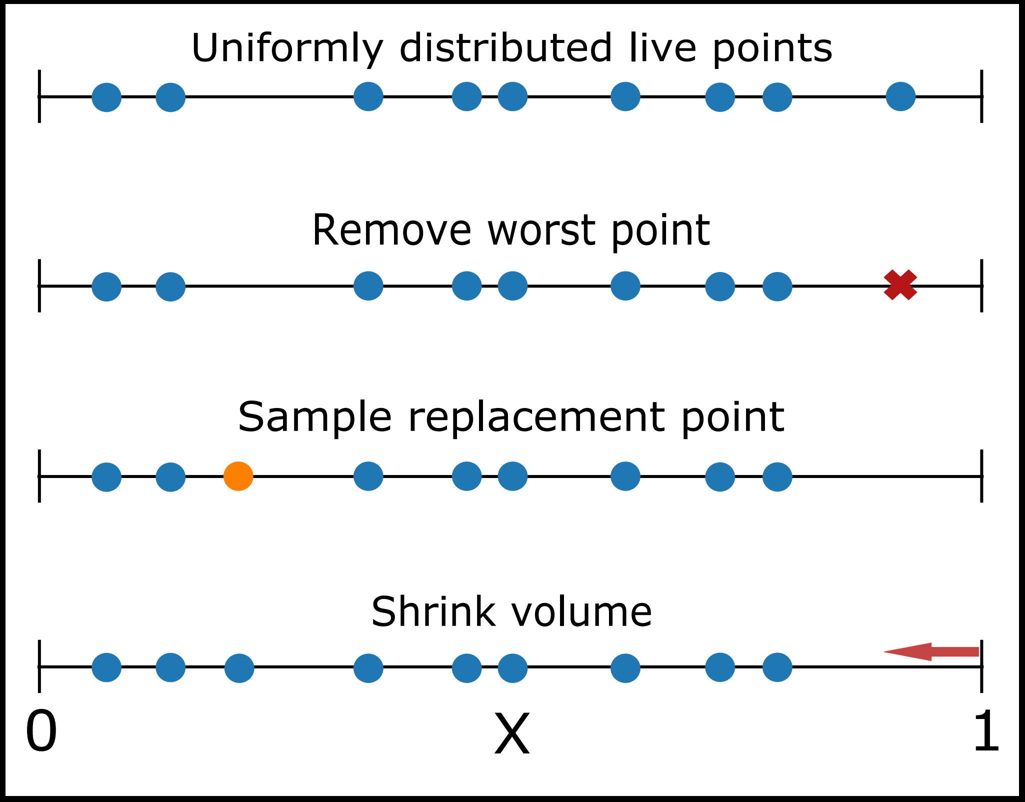
Affine-Invariant Ensemble Sampler

MCMC Methods

Nested Sampling

- Nested Sampling is specifically designed to calculate the evidence, also known as the marginal likelihood, in Bayesian statistics.
- **Core Concept:** Sequentially replaces the least likely samples with more probable ones from the prior distribution, compacting the sample space and focusing on the most relevant regions.
- **Process:** Starts with a random set of samples from the prior. At each iteration, the sample with the lowest likelihood is replaced with a new sample that has a higher likelihood.
- **Benefits:** Efficiently handles multi-modal distributions and provides not only evidence but also posterior samples.
- **Applications:** Extensively used in fields like astrophysics, statistical mechanics, and machine learning for model comparison and parameter estimation.
- **Key Takeaway:** Nested Sampling not only enhances the computation of Bayesian evidence but also offers a robust framework for exploring complex probabilistic models.

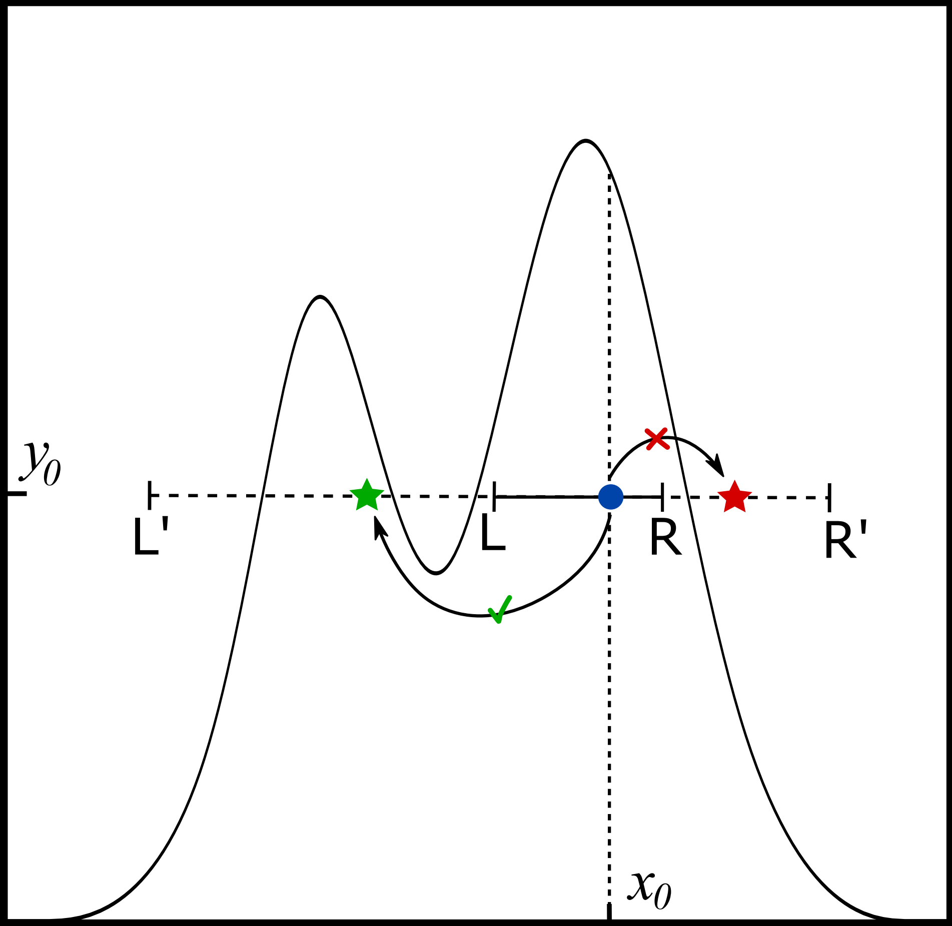




MCMC Methods

Slice Sampling

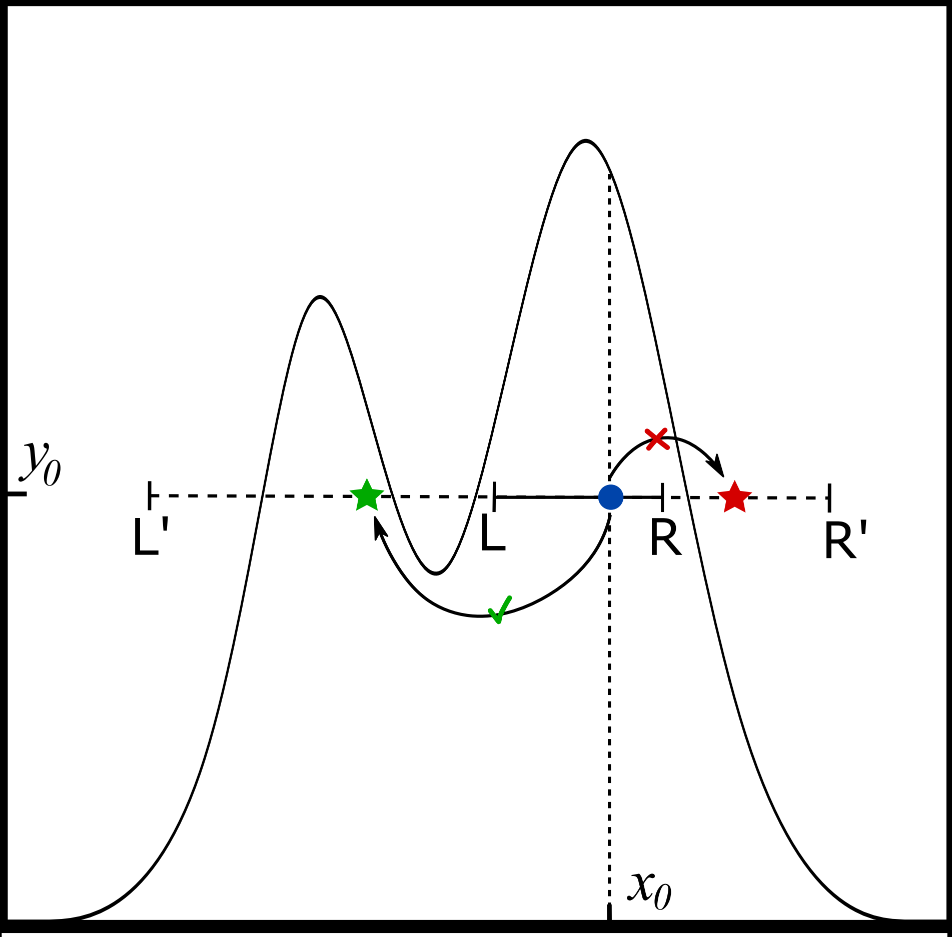
- Slice Sampling is an MCMC technique that improves sampling efficiency by focusing on relevant regions of the distribution.
- **Principle:** Samples are drawn by defining a 'slice' where the probability density is above a certain threshold, reducing dependence on the proposal distribution's tuning.
- **Process:** For a given current sample, a random threshold level is chosen below the density. A new sample is then drawn uniformly from the region where the density is above this threshold.
- **Advantages:** Automatically adapts to the shape of the distribution, requires less tuning of parameters, and can effectively handle multimodal distributions.
- **Applications:** Used in various fields requiring robust Bayesian inference, including statistics, machine learning, and astrophysics modelling.
- **Key Takeaway:** Slice Sampling offers a straightforward and adaptive approach to explore complex probability distributions, enhancing the efficiency and effectiveness of MCMC methods.

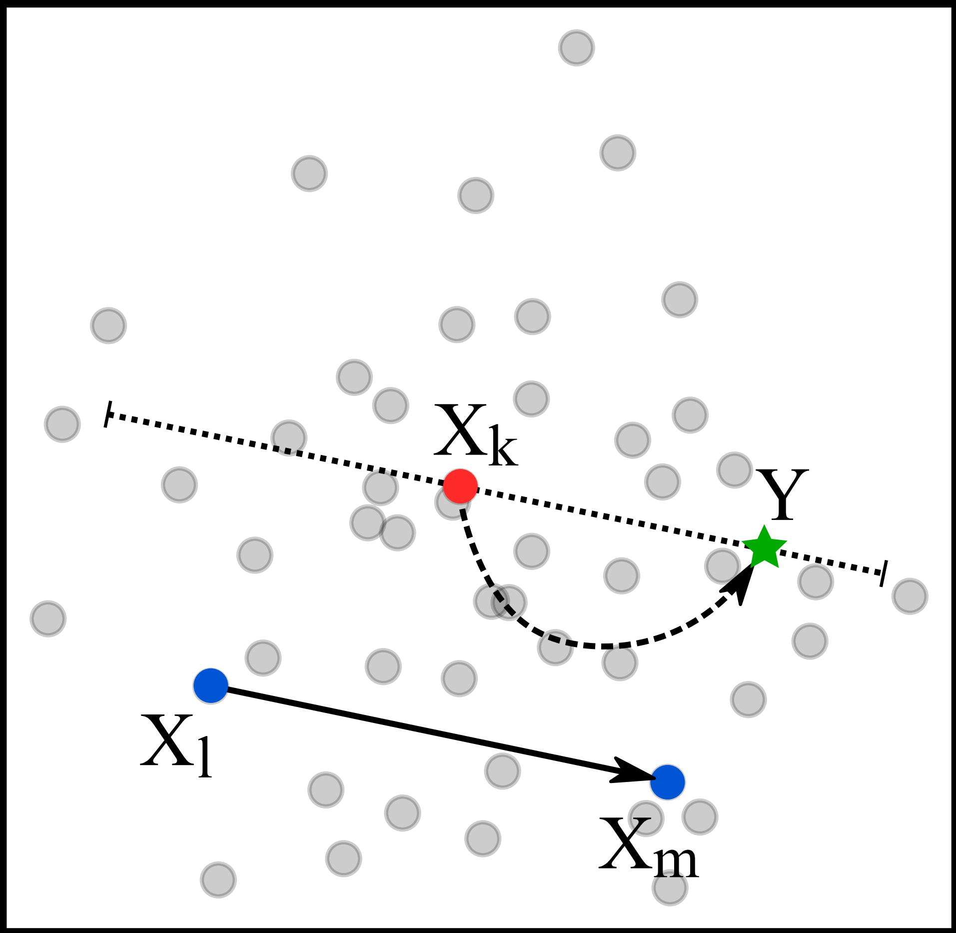


MCMC Methods

Ensemble Slice Sampling

- Ensemble Slice Sampling enhances traditional slice sampling by utilising multiple interacting samples to explore the state space more efficiently.
- **Concept Overview:** Combines the benefits of slice sampling with ensemble methods to prevent individual chains from becoming stuck in local modes.
- **Mechanism:** Involves a cooperative group of samples (or 'walkers') that adaptively adjust their step sizes based on their neighbours, promoting diversity and coverage of the probability distribution.
- **Advantages:** This method is particularly effective in dealing with correlated variables and complex posterior distributions.
- **Applications:** Useful in high-dimensional Bayesian inference tasks where traditional methods struggle with dependency and local optima.
- **Key Takeaway:** Ensemble Slice Sampling offers a powerful alternative for MCMC sampling, improving convergence times and the exploration of complex parameter spaces.





MCMC Methods

Comparison between methods

	emcee/AIES	emcee/DEMC	**zeus**
Cosmological inference			
efficiency^{-1}	12140	6750	**1320**
convergence cost	24×10^5	22×10^5	$\mathbf{6.6 \times 10^5}$
convergence fraction	7/40	14/40	**36/40**
Exoplanet inference			
efficiency^{-1}	1386	338	**47**
convergence cost	36.0×10^2	17.1×10^2	$\mathbf{4.8 \times 10^2}$
convergence fraction	23/40	29/40	**38/40**