



# MCMC SAMPLING

---

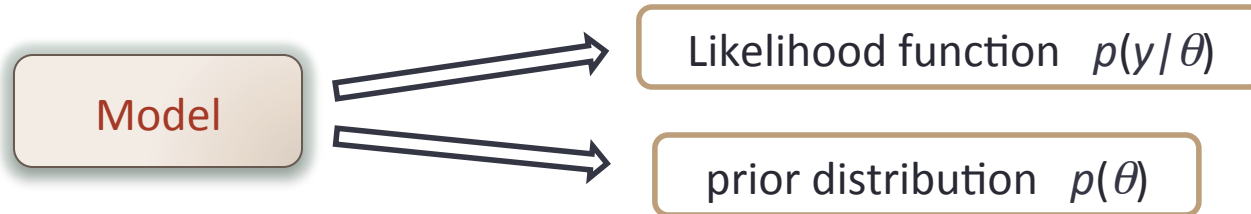
Sudhir Raman  
Translational Neuromodeling Unit,  
UZH & ETH



Recap

# Recap - Bayesian modelling

⇒ Formulation of a generative **model**



⇒ Observation of **data**

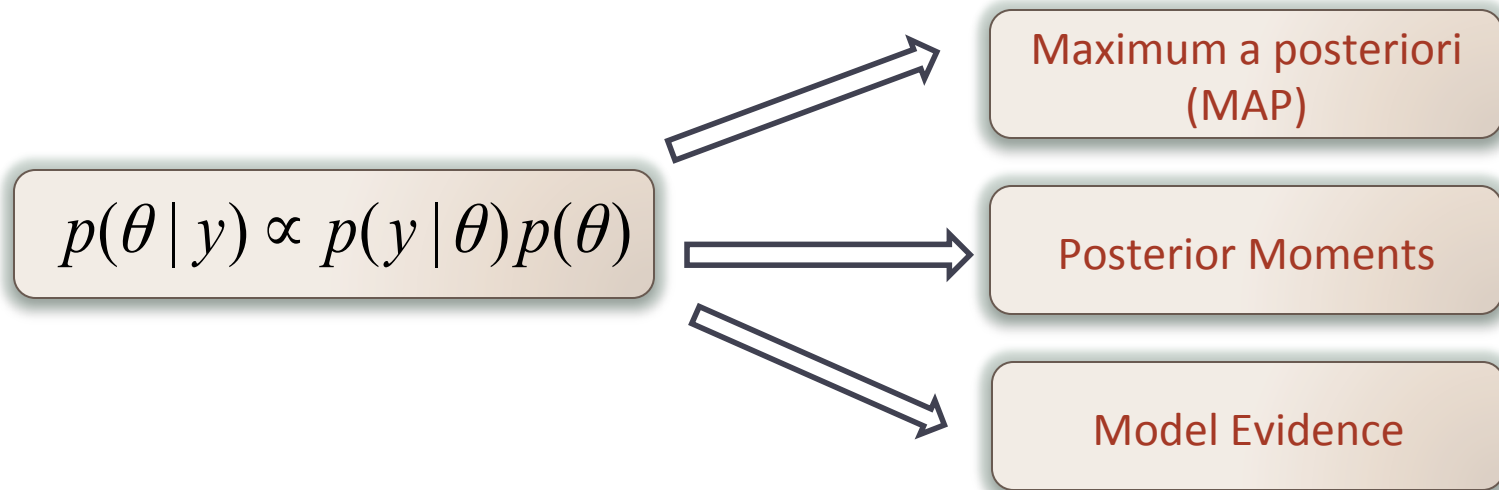


⇒ **Model Inversion** - **Update** of beliefs based upon observations, given a prior state of knowledge

$$P(\theta | y) = \frac{p(y | \theta) p(\theta)}{p(y)}$$



# Recap - Bayesian modelling

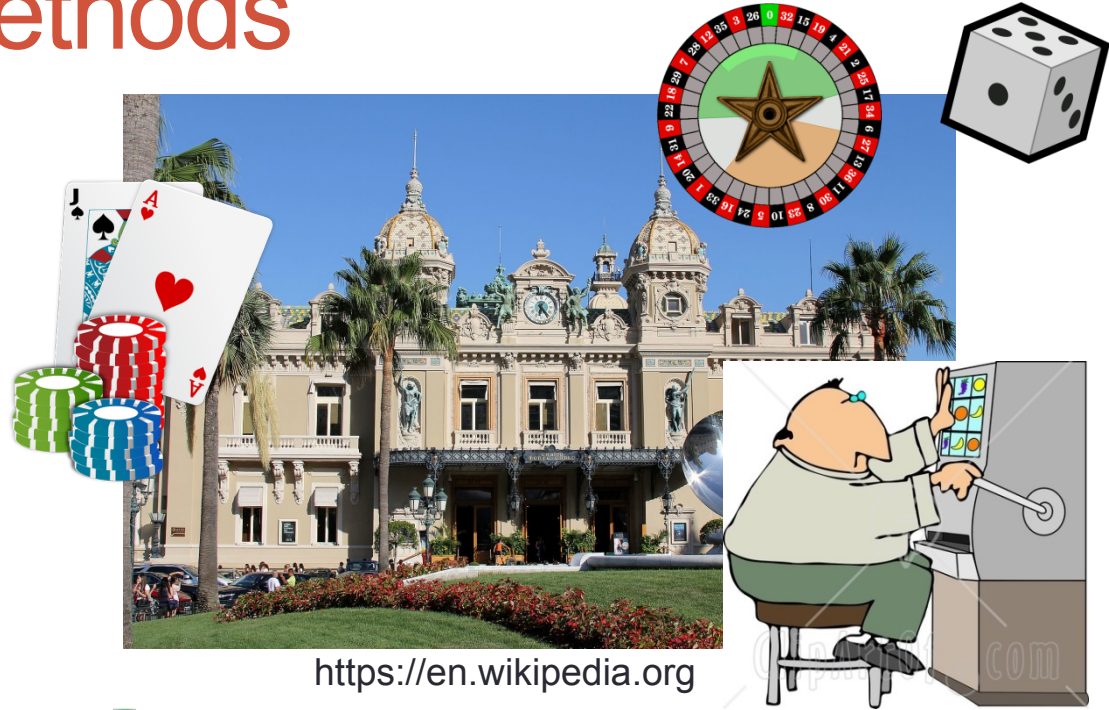


Deterministic approximation
<ul style="list-style-type: none"><li>• Variational Bayes</li></ul>

Stochastic approximation
<ul style="list-style-type: none"><li>• Sampling Methods</li></ul>

# Monte Carlo Methods

# Monte Carlo methods



<https://en.wikipedia.org>

Random Numbers

⇒ Expectation

⇒ Density

$$\int \mathbb{1}_A(\theta) f(\theta) p(\theta) d\theta$$



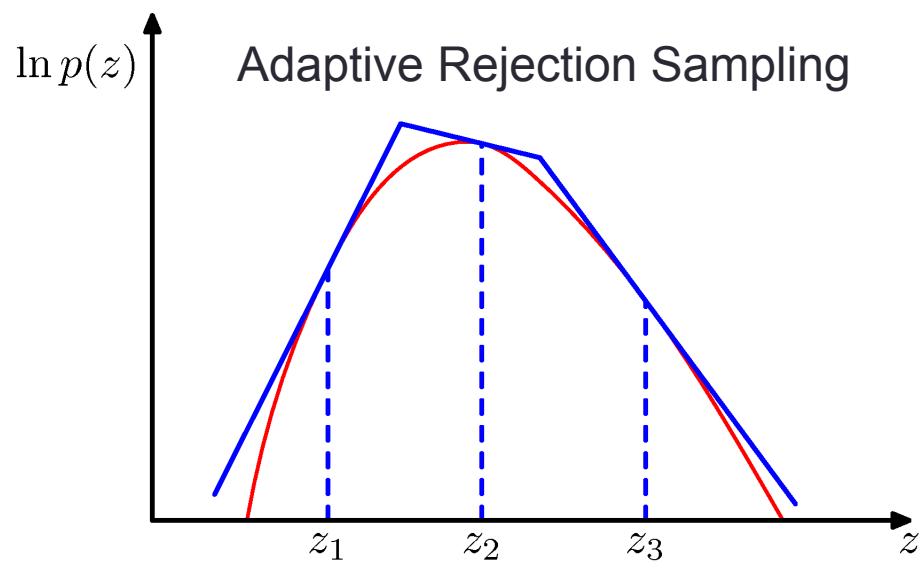
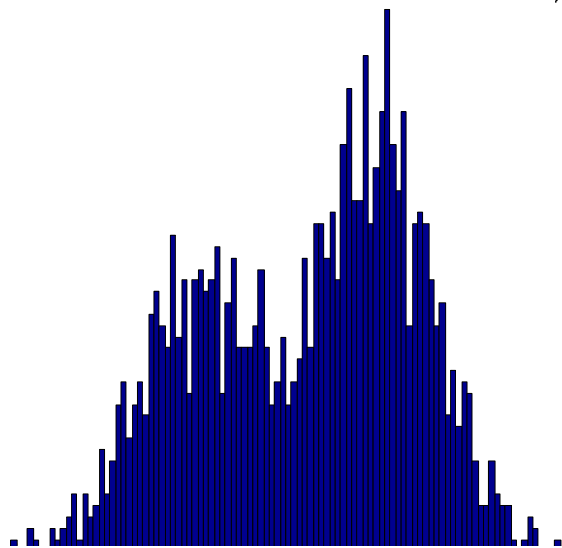
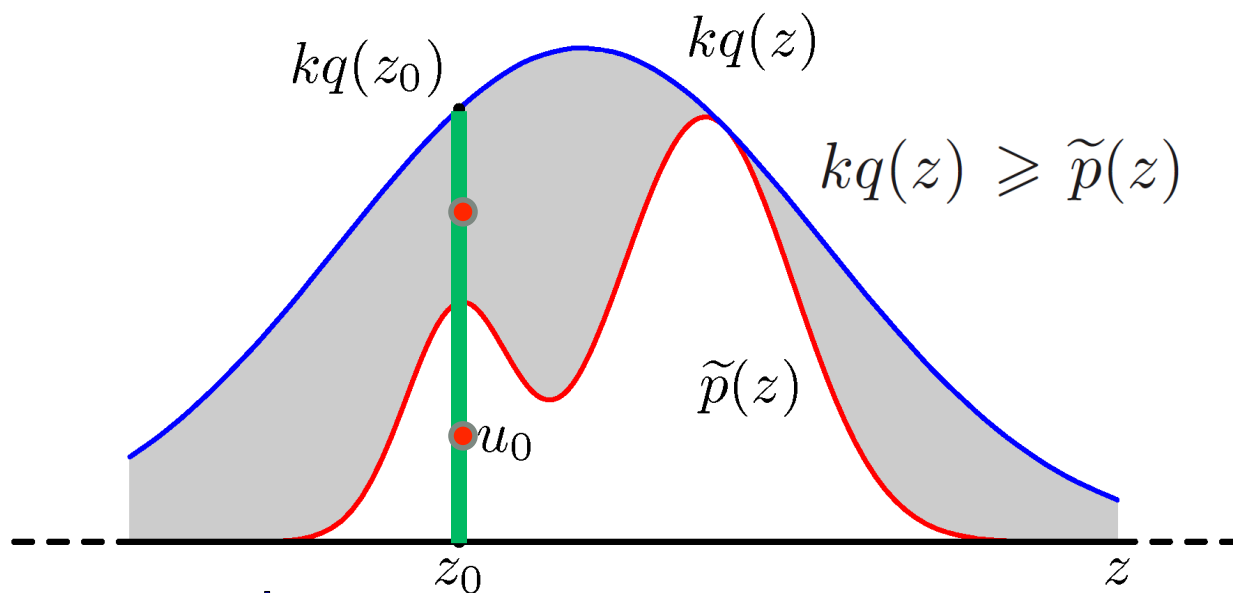
# Simple Monte Carlo

- What is the average height of people in this room ?



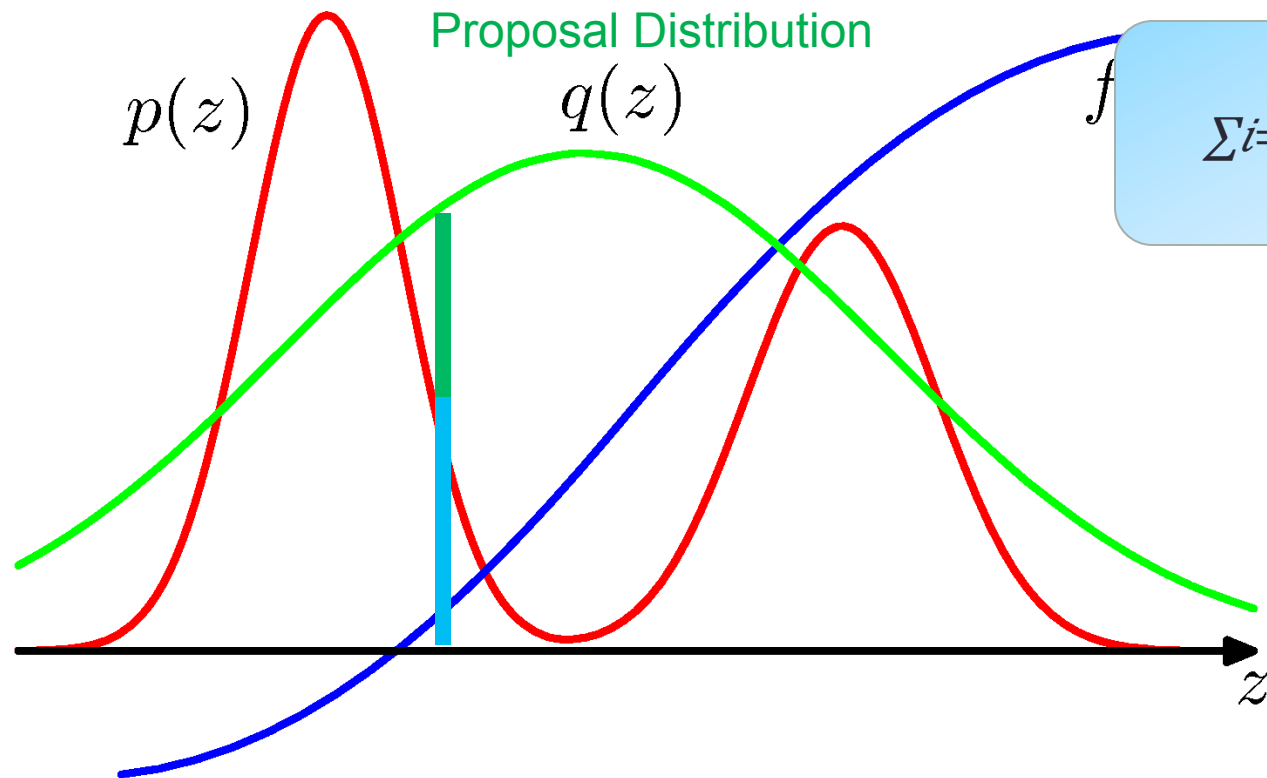
$$\frac{1}{N} \sum_{i=0}^{N-1} f(z_i)$$

# Rejection Sampling





# Importance Sampling



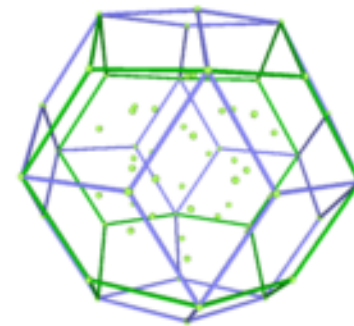
$$\int f(z)p(z)dz$$

$$\frac{1}{N} \sum_{i=0}^{N-1} f(z^{(i)})$$

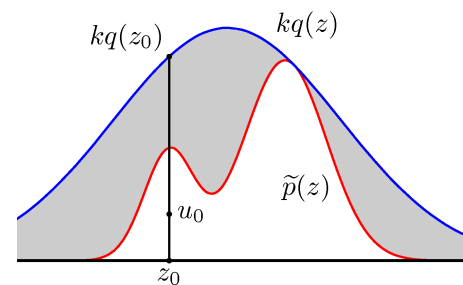
$$\mathbb{E}[f] \simeq \sum_{l=1}^L w_l f(\mathbf{z}^{(l)}) \quad w_l = \frac{\tilde{r}_l}{\sum_m \tilde{r}_m} = \frac{\tilde{p}(\mathbf{z}^{(l)})/q(\mathbf{z}^{(l)})}{\sum_m \tilde{p}(\mathbf{z}^{(m)})/q(\mathbf{z}^{(m)})}.$$

# Issues

Scaling to high dimensions



Choice of proposal distribution



# Markov Chain Monte Carlo

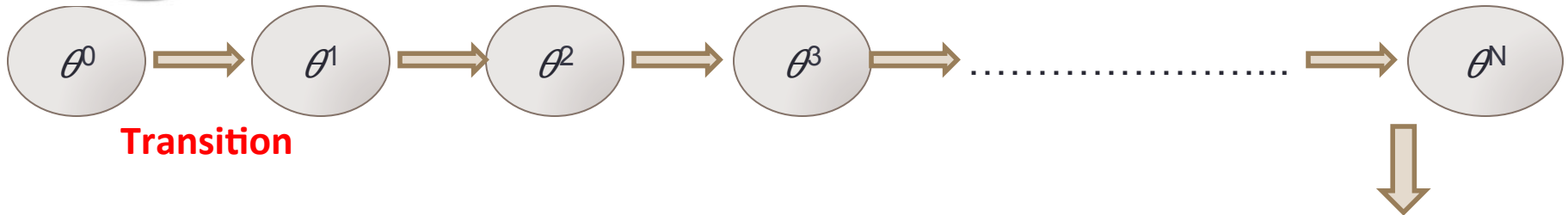
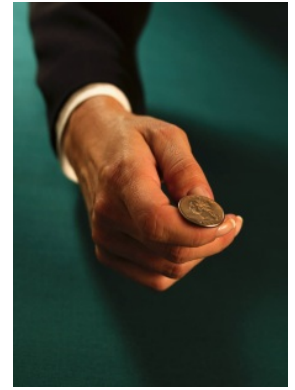
# Markov Chain Monte Carlo(MCMC) sampling



**Andrei Markov**  
(1856 – 1922)  
Russian  
mathematician

Proposal  
distribution

$$q(\theta)$$



**Transition**

- ▣ A general framework for sampling from a large class of distributions
- ▣ Scales well with dimensionality of sample space
- ▣ Asymptotically convergent

Posterior  
distribution

$$p(\theta|y)$$

# Markov chain properties

- Transition probabilities – homogeneous

$$p(\theta_{t+1} \mid \theta_1, \dots, \theta_t) = p(\theta_{t+1} \mid \theta_t) = T(\theta_t, \theta_{t+1})$$

- Invariance

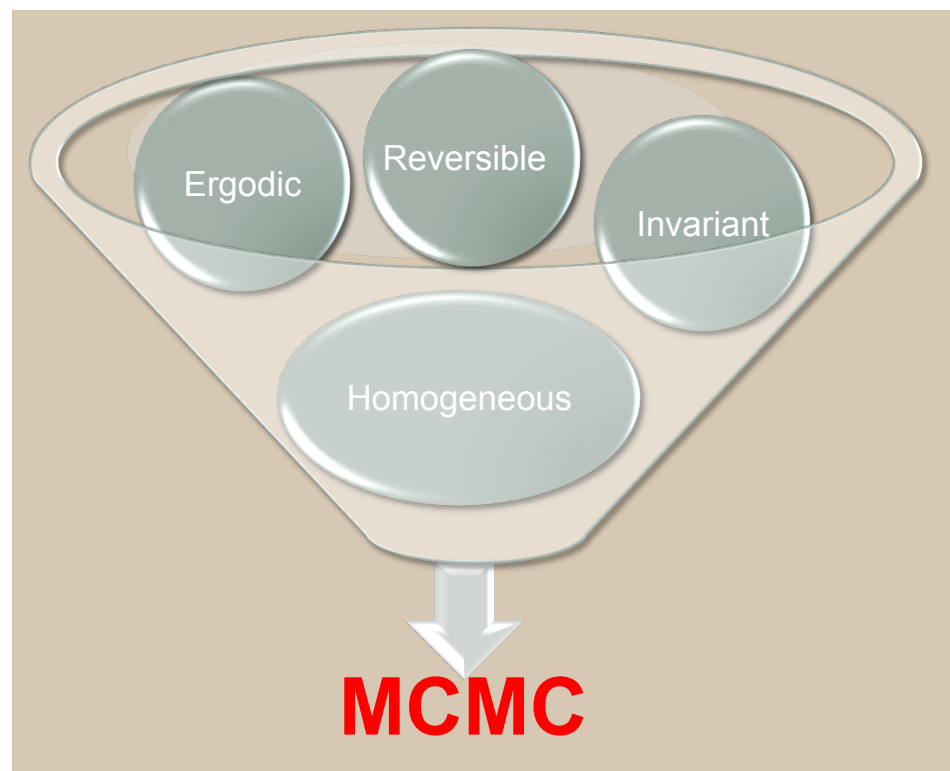
$$p^*(\theta) = \sum_{\theta'} T(\theta', \theta) p^*(\theta')$$

- Detailed Balance

$$T(\theta, \theta') p^*(\theta) = T(\theta', \theta) p^*(\theta')$$

- Ergodicity

$$p^*(\theta) = \lim_{n \rightarrow \infty} p(\theta_n) \quad \forall p(\theta_0)$$



# Metropolis-Hastings Algorithm

$$p(\theta|y)$$

- Initialize  $\theta$  at step 1 - for example, sample from prior
- At step  $t$ , sample from the proposal distribution:

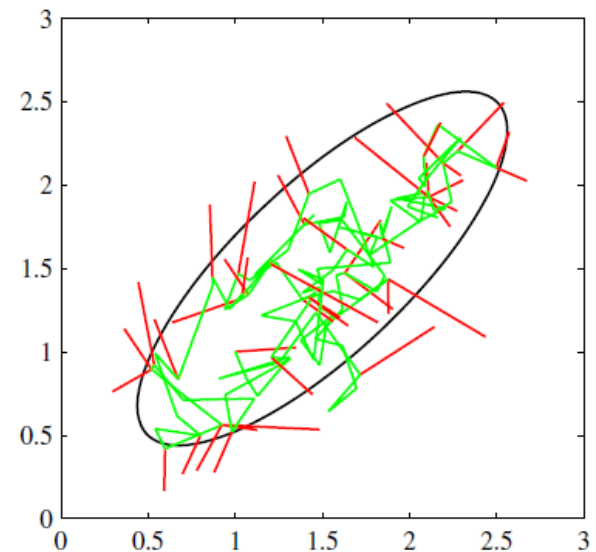
$$\theta^* \sim q(\theta^* | \theta^t)$$

- Accept with probability:

$$A(\theta^*, \theta^t) \sim \min(1, p(\theta^* | y) q(\theta^t | \theta^*) / p(\theta^t | y) q(\theta^* | \theta^t))$$

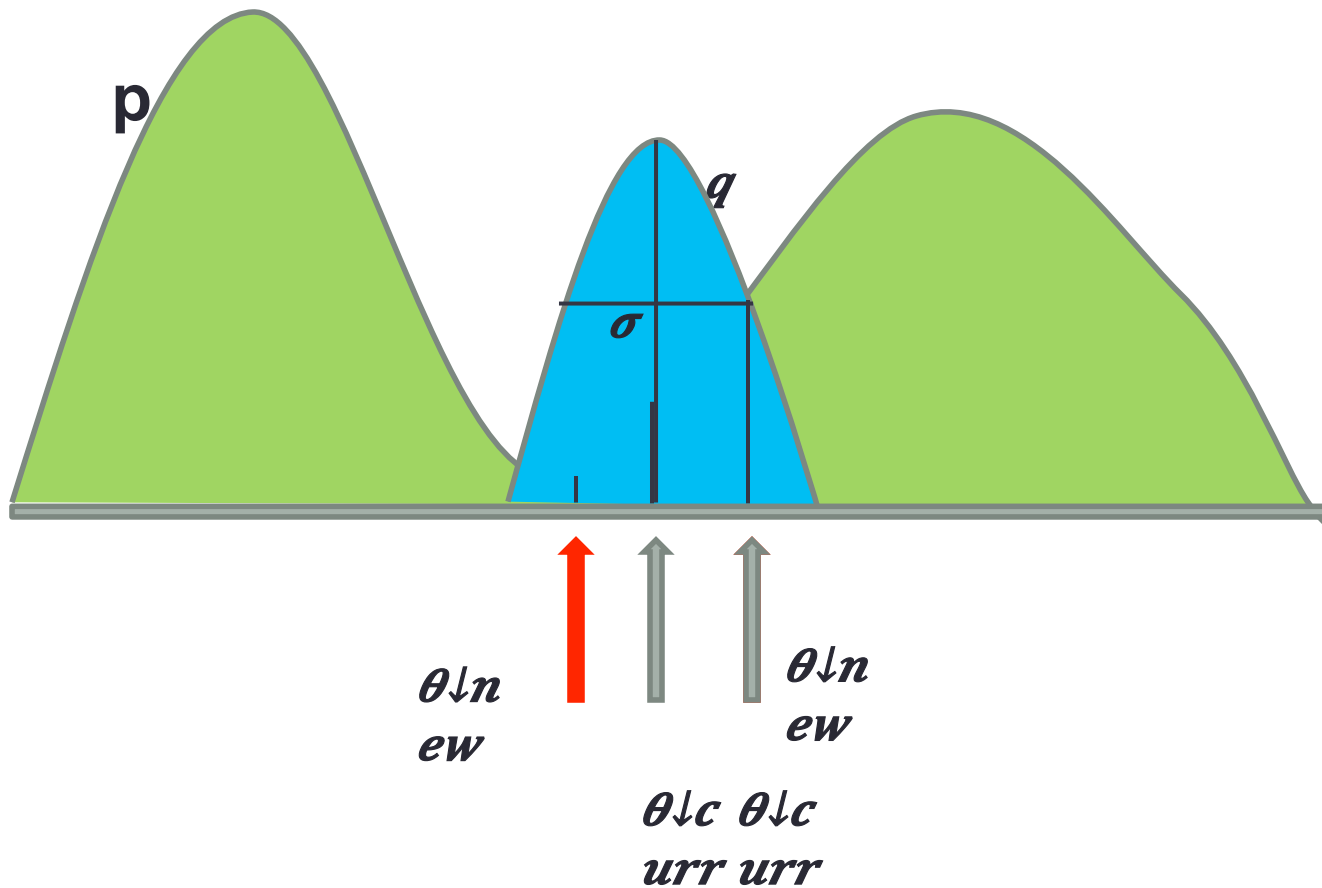
- Metropolis – Symmetric proposal distribution

$$A(\theta^*, \theta^t) \sim \min(1, p(\theta^* | y) / p(\theta^t | y))$$



Bishop (2006) PRML, p. 539

# Visual Example



# Gibbs Sampling Algorithm

$$p(\boldsymbol{\theta}) = p(\theta_1, \theta_2, \dots, \theta_n)$$

- Special case of Metropolis Hastings
- At step  $t$ , sample from the conditional distribution:

$$\begin{aligned} \theta_1^{(t+1)} &\sim p(\theta_1 | \theta_2^{(t)}, \dots, \theta_n^{(t)}) \\ \theta_2^{(t+1)} &\sim p(\theta_2 | \theta_1^{(t+1)}, \dots, \theta_n^{(t)}) \\ &\vdots \\ \theta_n^{(t+1)} &\sim p(\theta_n | \theta_1^{(t+1)}, \dots, \theta_{n-1}^{(t+1)}) \end{aligned}$$

- Acceptance probability = 1
- Blocked Sampling

$$\theta_1^{(t+1)}, \theta_2^{(t+1)} \sim p(\theta_1, \theta_2 | \theta_3^{(t)}, \dots, \theta_n^{(t)})$$



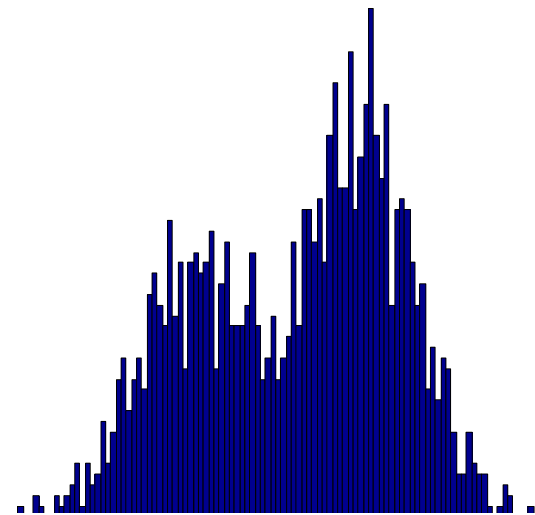


# Posterior analysis from MCMC



## Obtain independent samples:

- Generate samples based on MCMC sampling.
- Discard initial “burn-in” period samples to remove dependence on initialization.
- Thinning- select every  $m^{\text{th}}$  sample to reduce correlation .
- Inspect sample statistics (e.g., histogram, sample quantiles, ...)

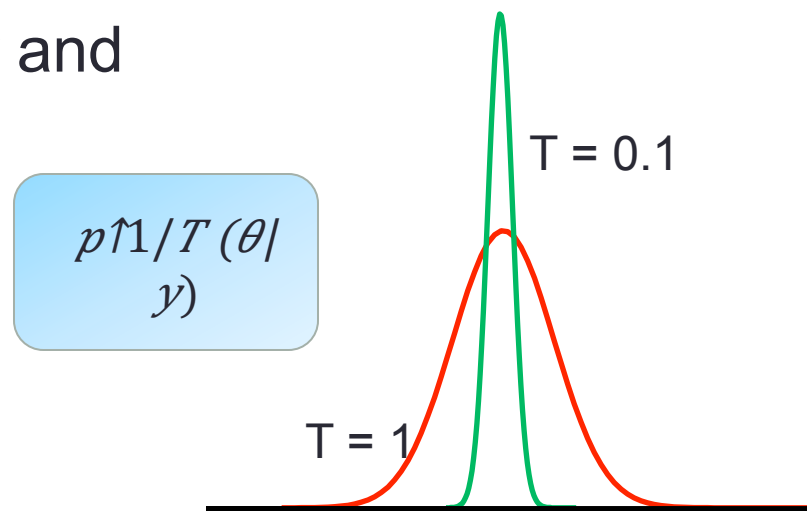


# MAP estimate via Simulated Annealing

- Add a temperature parameter and schedule to update it

- Algorithm

- Set  $T = 1$
- Until convergence
  - For every  $K$  iterations sample from:
  - Reduce  $T$





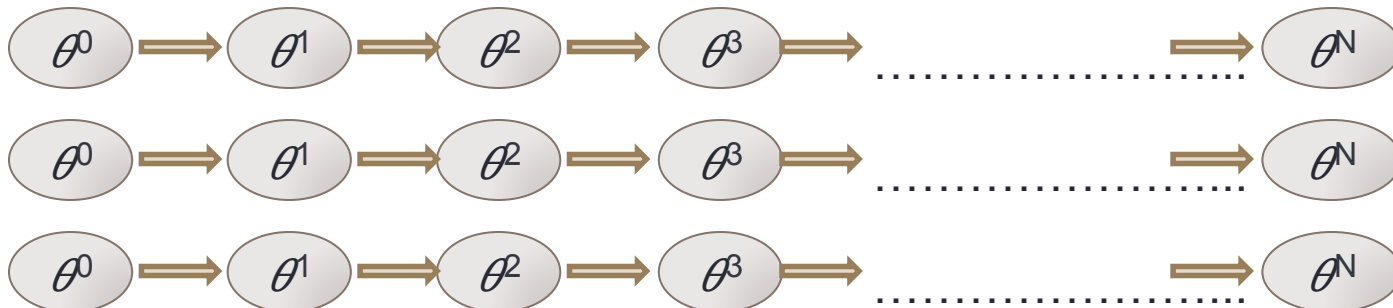
# Convergence Analysis

**Are We  
THERE Yet?**

- Single chain methods
  - Geweke (1992)
  - Raftery-Lewis (1992)



- Multi-chain methods
  - Gelman-Rubin – (1992)
    - Potential Scale Reduction factor



# Model Evidence

# Model evidence using MCMC

- Importance Sampling

$$p(D | M) = \frac{E_g \left[ \frac{p(D|\theta, M)p(\theta|M)}{g(\theta)} \right]}{E_g \left[ \frac{p(\theta|M)}{g(\theta)} \right]},$$

---

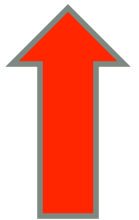
- Prior arithmetic mean

$$\widehat{f(Y)} = \frac{1}{M} \sum_{m=1}^M p(Y|\theta_m)$$



- Posterior harmonic mean

$$\widehat{f(Y)} = \frac{1}{\frac{1}{M} \sum_{m=1}^M \frac{1}{p(\mathbf{Y}|\bullet)}},$$



# Thermodynamic Integration

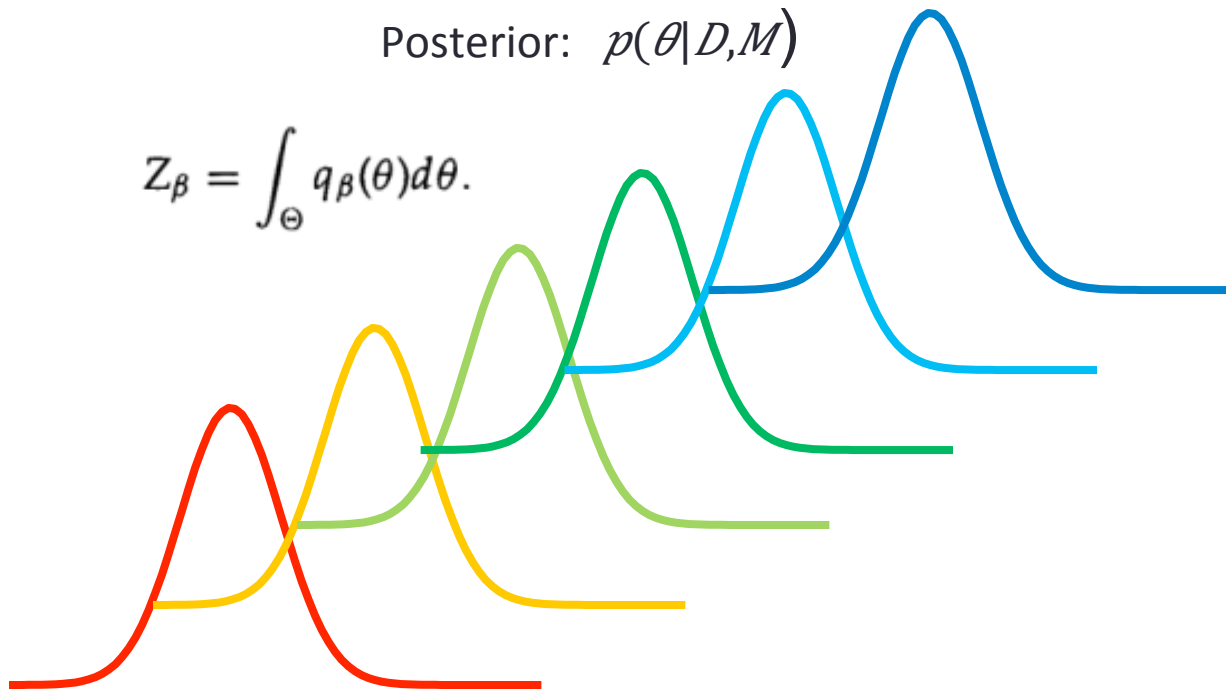
- Use multiple distributions

$$q_{\beta}(\theta) = p(D | \theta, M)^{\beta} p(\theta | M).$$

Posterior:  $p(\theta | D, M)$

$$Z_{\beta} = \int_{\Theta} q_{\beta}(\theta) d\theta.$$

Prior:  $p(\theta | M)$



$$p_{\beta}(\theta) = \frac{1}{Z_{\beta}} q_{\beta}(\theta), \quad (15)$$

$$Z_{\beta} = \int_{\Theta} q_{\beta}(\theta) d\theta. \quad (16)$$

When  $\beta$  tends to 0 (resp. 1),  $p_{\beta}$  converges pointwise to  $p_0$  (resp.  $p_1$ ), and  $Z_{\beta}$  to  $Z_0$  (resp.  $Z_1$ ).

Taking the derivative of  $\ln Z_{\beta}$  with respect to  $\beta$ :

$$\frac{\partial \ln Z_{\beta}}{\partial \beta} = \frac{1}{Z_{\beta}} \frac{\partial Z_{\beta}}{\partial \beta} \quad (17)$$

$$= \frac{1}{Z_{\beta}} \frac{\partial}{\partial \beta} \int_{\Theta} q_{\beta}(\theta) d\theta \quad (18)$$

$$= \frac{1}{Z_{\beta}} \int_{\Theta} \frac{\partial q_{\beta}(\theta)}{\partial \beta} d\theta \quad (19)$$

$$= \int_{\Theta} \frac{1}{q_{\beta}(\theta)} \frac{\partial q_{\beta}(\theta)}{\partial \beta} \frac{q_{\beta}(\theta)}{Z_{\beta}} d\theta \quad (20)$$

$$= \int_{\Theta} \frac{\partial \ln q_{\beta}(\theta)}{\partial \beta} p_{\beta}(\theta) d\theta \quad (21)$$

$$= E_{\beta} \left[ \frac{\partial \ln q_{\beta}(\theta)}{\partial \beta} \right], \quad (22)$$

$p_{\beta}$ . Defining the potential

$$U(\theta) = \frac{\partial \ln q_{\beta}(\theta)}{\partial \beta}, \quad (23)$$

one has thus the first moment identity:

$$\frac{\partial \ln Z_{\beta}}{\partial \beta} = E_{\beta}[U]. \quad (24)$$

Integrating over  $[0, 1]$  yields the log-ratio one is looking for:

$$\mu = \ln Z_1 - \ln Z_0 \quad (25)$$

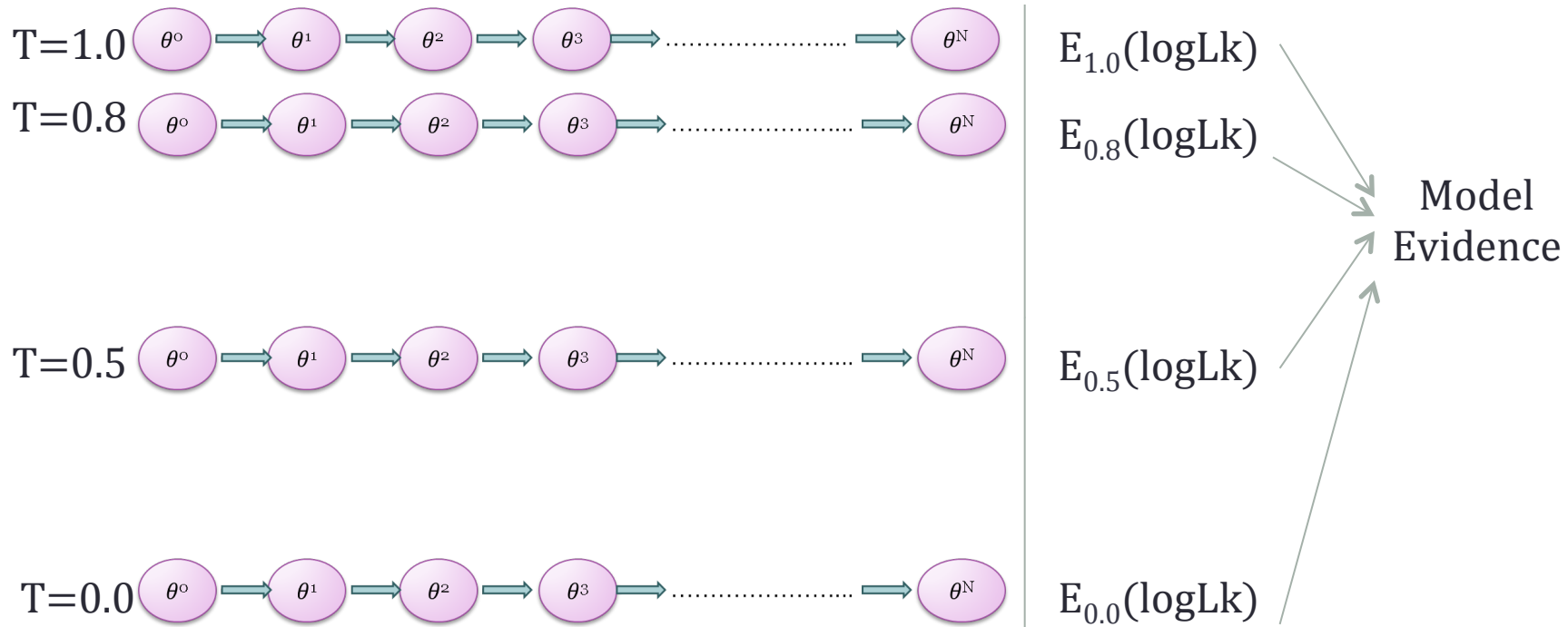
$$= \int_0^1 \frac{\partial \ln Z_{\beta}}{\partial \beta} d\beta \quad (26)$$

$$= \int_0^1 E_{\beta}[U] d\beta. \quad (27)$$

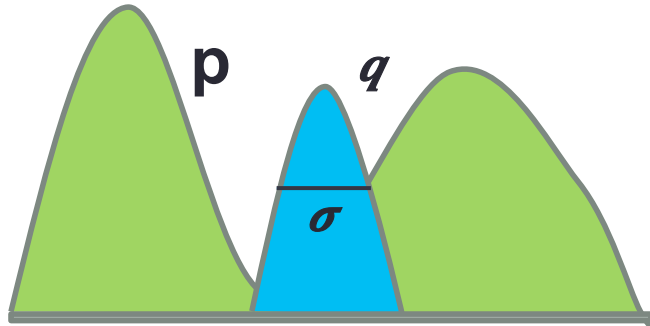
# Thermodynamic Integration

- Path Sampling (Thermodynamic Integration)

$$q_{\beta}(\theta) = p(D | \theta, M)^{\beta} p(\theta | M).$$

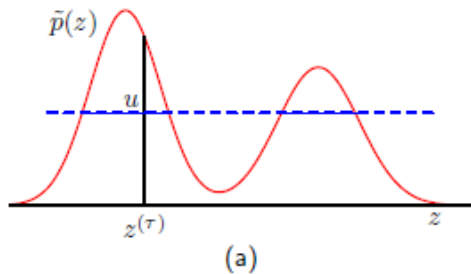
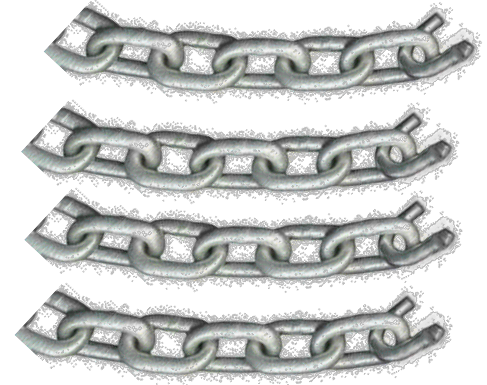


# Other MCMC variants



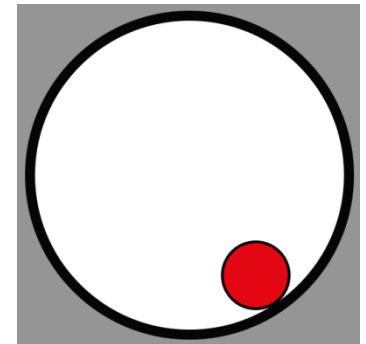
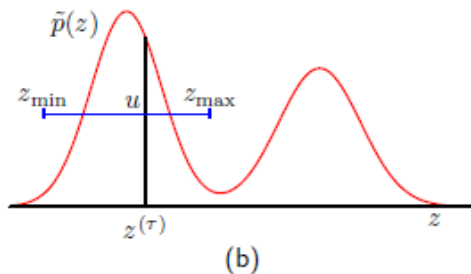
Adaptive  
MCMC

Population  
MCMC



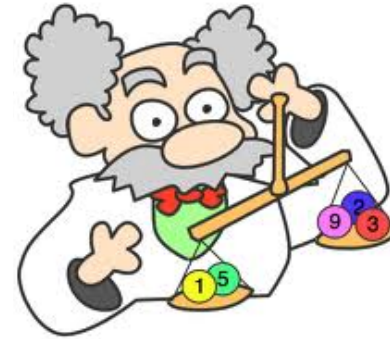
Slice  
Sampling





Hybrid  
Monte Carlo





# Comparison



Criteria	Variational Bayes	MCMC Sampling
Method	Find a proxy posterior using a hypothesis class	Collect samples
Speed		
Model Evidence	Free	Computationally Expensive
Convergence	Local minima	Convergence diagnostics required
Update Equations		

**Translational Neuromodeling Unit**  
Institute for Biomedical Engineering, University of Zurich & ETH Zurich

official website [uzh/ethz](#)

HOME NEWS RESEARCH PROGRAMME TEAM HISTORY OF THE TNU CONTACT FORSCHUNGSAMBULANZ

TEACHING SPM COURSE MSc PROJECTS **SOFTWARE** PUBLICATIONS OPEN POSITIONS

## TAPAS - TNU Algorithms for Psychiatry-Advancing Science

**TAPAS** is a collection of algorithms and software tools that are developed by the Translational Neuromodeling Unit (TNU) at Zurich. These tools have been developed to support translational neuroscience, particularly concerning the application of neuroimaging and computational modeling to research questions in psychiatry and neurology. Problems that can be addressed by tools in TAPAS presently include:

- **PhysIO**: Physiological noise correction of fMRI data.
- **HGF**: Hierarchical Gaussian Filtering (Bayesian inference on computational processes from observed behaviour).
- **MTCF**: Mixed-effects inference on classification performance.
- **VBLM**: Variational Bayes for linear regression models.
- **mpdcm**: Massively parallel DCM (efficient integration of dynamical systems in DCM).

TAPAS is written in **MATLAB** and distributed as open source code under the GNU General Public License (**GPL, Version 3**).

### Download

The software can be downloaded after completing a **registration form**. The information will be used for generating statistics regarding the usage of this software. Submitting the form will take you to the page where the latest version can be downloaded. Older versions of TAPAS can be downloaded [here](#).

### Mailing List

For discussion/queries/suggestions, please subscribe to the mailing list at [TAPAS@sympa.ethz.ch](mailto:TAPAS@sympa.ethz.ch). Users without an ETH ID/password can register with their email-ID by clicking the **"Subscribe"** link on the left side of the mailing list page.

We are in a continuous process of improving our software, hence it would be helpful for us to know about any bugs that you encounter in the current version. The same mailing list can be used for reporting any bugs in the software. For ease of management, please mention the specific operation/toolbox, where the error was found, in the subject line, along with details/snapshots of the error message.

### Documentation

Detailed description of the software can be found in the [Documentation](#) section.

### Data

Sample code and datasets can be found in the [Data](#) section.

### Publications

The publications associated with the software are listed in the [Publications](#) section.

**Tags:** autism Bayesian Brains CNC course DCM Dopar EEG fag fMRI HGF Impulsivity Inference Magnetic Field Monitoring Matched-filter Acquisition neuropharmacology Physiological Noise Resting-state fMRI schizophrenia Social decision-making Statistics TAPAS teaching Variational Bayes videos

**SEMINARS & EVENTS**  
Zurich Computational Psychiatry Meeting 2014  
*Psychiatry under the Lens of Algorithms*  
May 19/20, 2014  
[» more information](#)

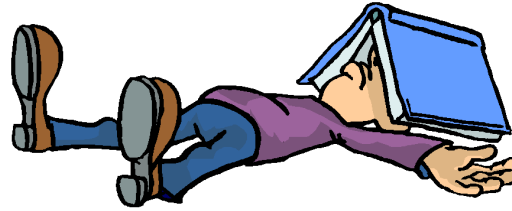
**INTERNAL**  
[» Log in](#)  
[» TNU Wiki](#)

- mpdcm – GPU based Population MCMC for DCM
- <http://www.translationalneuromodeling.org/tapas/>

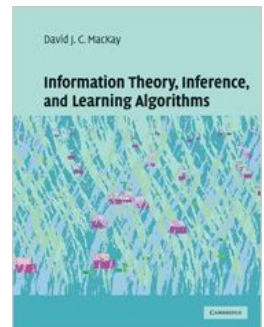
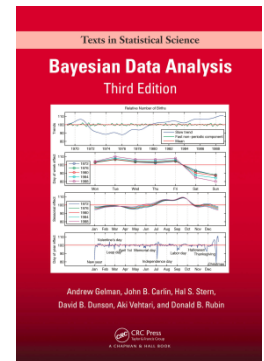
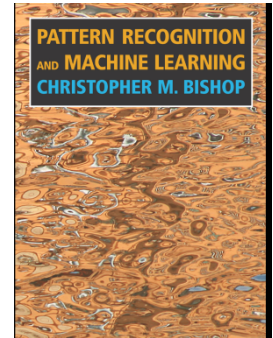
# Conclusion

- Simple, Asymptotically Exact
- Computationally Expensive
- Profiling, Parallelizing
- Testing, Testing, Testing, Testing

# References



- Pattern Recognition and Machine Learning (2006). Christopher M. Bishop
- Information theory, pattern recognition and learning algorithms. David McKay
- Videolectures.net – Bayesian inference, MCMC , Variational Bayes
- Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition (Chapman & Hall/CRC Texts in Statistical Science) Dani Gamerman, Hedibert F. Lopes.
- N. Lartillot and H. Philippe, “Computing bayes factors using thermodynamic integration,” Systematic Biology, vol. 55, no. 2, pp. 195–207, 2006.
- Andrew Gelman, Donald B. Rubin. Inference from iterative simulation using multiple sequences Stat. Sci., 7 (4) (1992), pp. 457–472
- J. Geweke. Evaluating the accuracy of sampling-based approaches to calculating posterior moments Bayesian Statistics 4, (1992), pp. 169–193
- Adrian E. Raftery, Steven Lewis. How many iterations in the Gibbs sampler? Bayesian Statistics 4, Oxford University Press (1992), pp. 763–773



Thank You