Optimization for Hamiltonian Monte Carlo

Aisaku Arakawa, Motohide Nishio, Masaaki Taniguchi, Satoshi Mikawa

Institute of Livestock and Grassland Science, NARO, Japan

Hamiltonian (Hybrid) Monte Carlo (HMC)

One of Markov Chain Monte Carlo methods
STAN uses the HMC algorithm
HMC works when conditional distributions are unknown,
→ a Metropolis-Hastings method

The HMC has one weak point in the process of sampling

Our Objectives

To apply Hamiltonian Monte Carlo methods to a univariate gaussian trait
To reveal the property of the Hamiltonian Monte Carlo methods

Belief introduction of Hamiltonian Monte Carlo

HMC follows the rule of **Hamilton dynamics**

$$H(\boldsymbol{\theta}, \mathbf{p}) \quad = \quad U(\boldsymbol{\theta}) \quad + \quad K(\mathbf{p})$$

Hamiltonian Potential Energy Kinetic Energy $\frac{1}{2}\mathbf{p}'\mathbf{M}^{-1}\mathbf{p}$

Hamiltonian on Bayesian inference

Posterior likelihood Priors

$$p(\mathbf{b},\mathbf{a},\sigma_a^2,\sigma_e^2|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{b},\mathbf{a},\sigma_a^2,\sigma_e^2)p(\mathbf{b},\mathbf{a},\sigma_a^2,\sigma_e^2) = p(\boldsymbol{\theta})$$

$$p(\boldsymbol{\theta},\mathbf{p}) = p(\boldsymbol{\theta}|\mathbf{y}) \times p(\mathbf{p})$$

Bayesian posterior

$p(\mathbf{p})$ **Independent** Arbitrary Parameter (vector)

= multivariate normal distribution

$$p(\mathbf{p}) = N(0,\mathbf{M}) \propto \exp\left(-\frac{1}{2}\mathbf{p}'\mathbf{M}^{-1}\mathbf{p}\right)$$

$$= \exp[\log p(\boldsymbol{\theta}|\mathbf{y}) + \log p(\mathbf{p})] \propto \exp\left[\log p(\boldsymbol{\theta}) - \frac{1}{2}\mathbf{p}'\mathbf{M}^{-1}\mathbf{p}\right]$$

$$H(\boldsymbol{\theta},\mathbf{p}) = -\log p(\boldsymbol{\theta}) + \frac{1}{2}\mathbf{p}'\mathbf{M}^{-1}\mathbf{p}$$

How to sample by using the HMC method

Hamilton's equations
(Differential equations)

$$\frac{d\boldsymbol{\theta}}{dt} = \frac{\partial H}{\partial \mathbf{p}} = \frac{dK}{d\mathbf{p}}$$

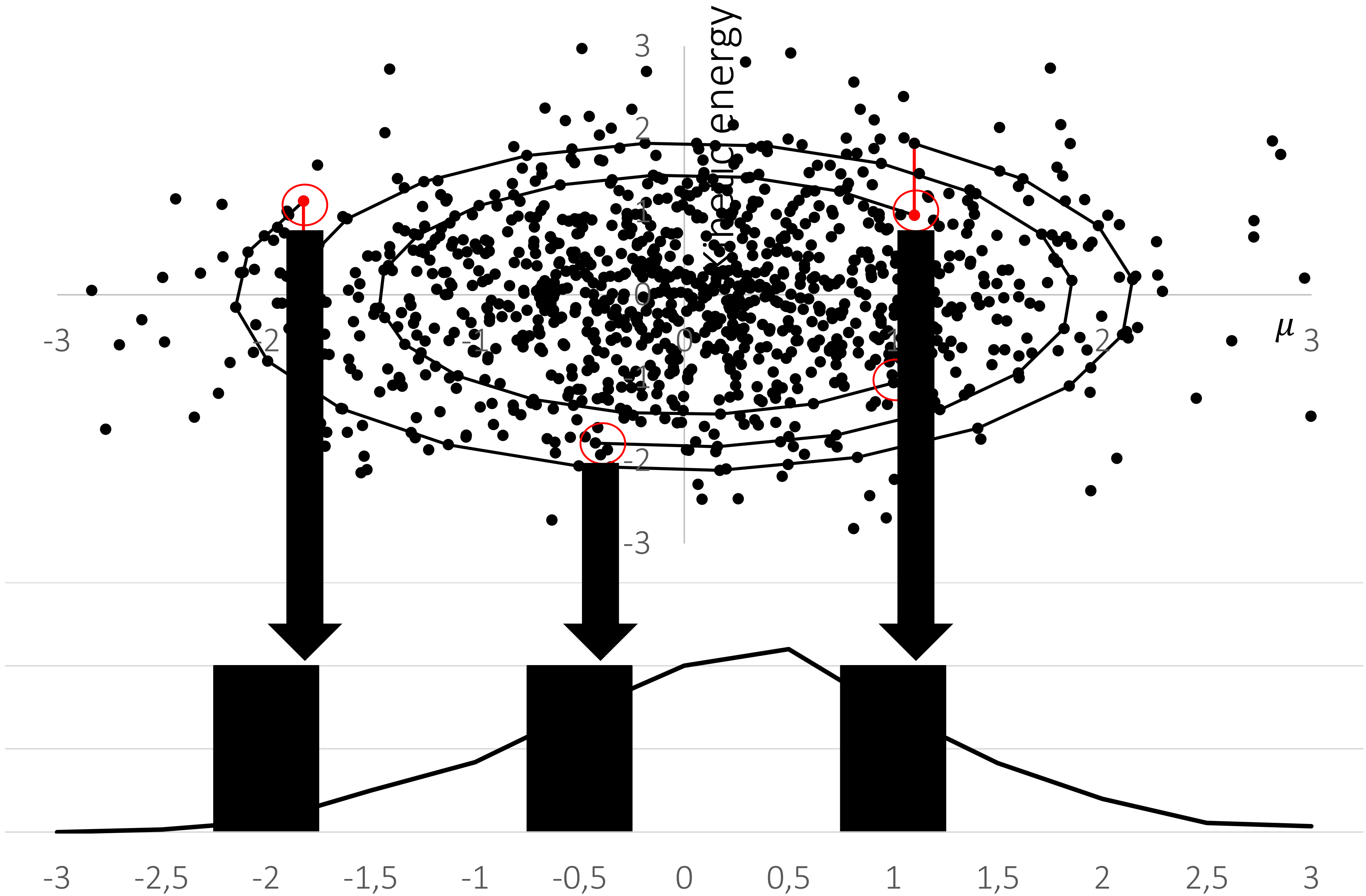
$$\frac{d\mathbf{p}}{dt} = -\frac{\partial H}{\partial \boldsymbol{\theta}} = -\frac{dU}{d\boldsymbol{\theta}}$$

t: time
$\boldsymbol{\theta}$: estimates
$\mathbf{p}$: kinetic energy

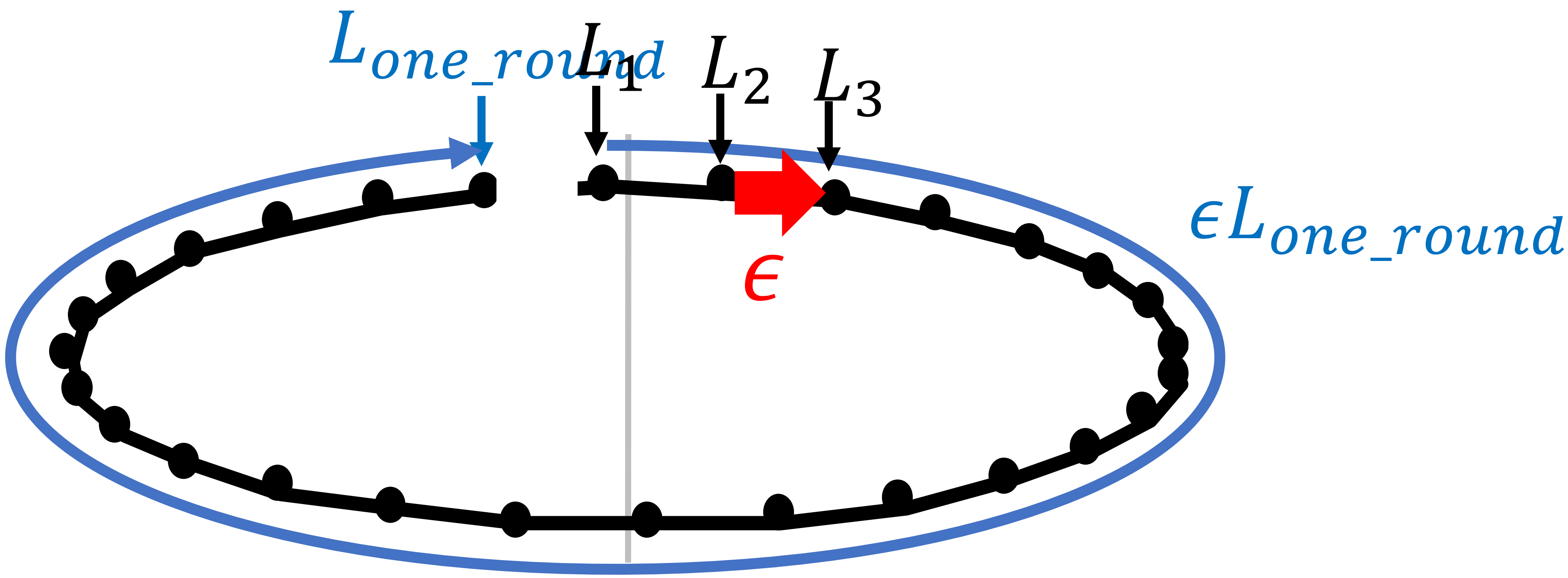
⬇

Leapfrog (Path) Integration

How to sample by using the HMC method ($N(0,1)$)



Problem on the leapfrog integration



L : Number of step on the leapfrog integration

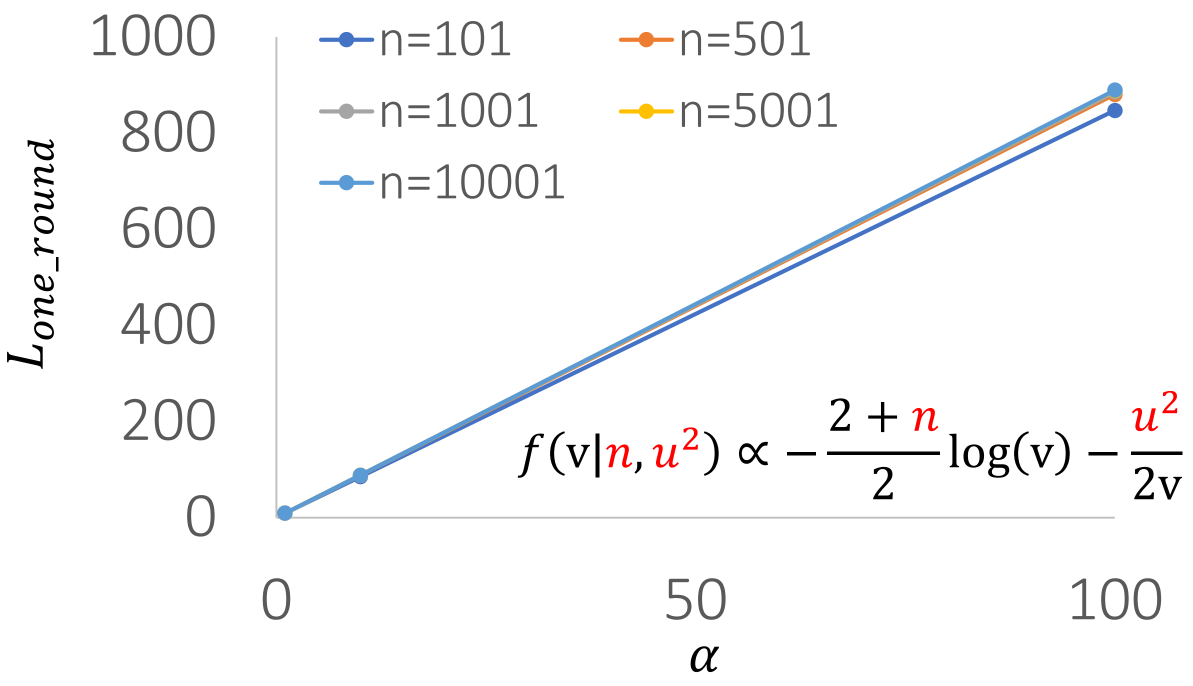
ϵ: Step size

L_{one_round}: The number iteration for one round

ϵL_{one_round}: Length of the trajectory

Tuning of Leapfrog step ($\epsilon = \sqrt{Var[\mathrm{v}]}/\alpha$)

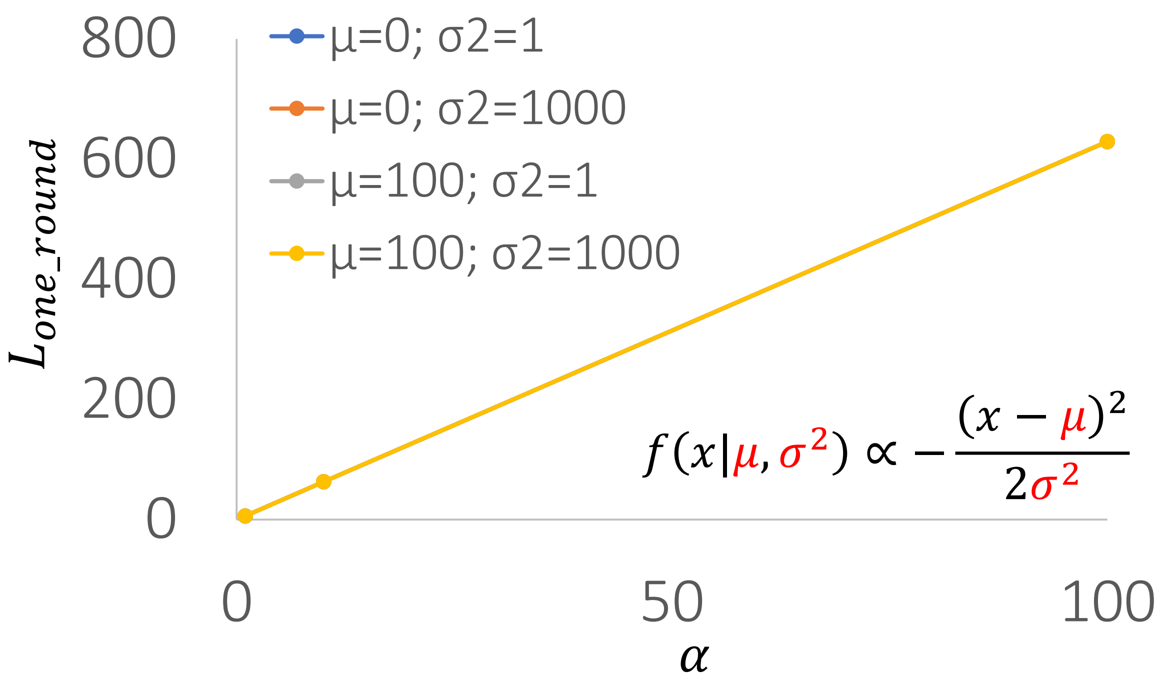
Inverse chi-square distribution



$$\epsilon L_{one\_round} \cong \sqrt{Var[\mathrm{v}]}/0.112$$

$$\epsilon \cong \sqrt{Var[\mathrm{v}]} \Big/ 0.112 \times L_{one\_round}$$

Normal distribution



$$\epsilon L_{one\_round} \cong \sqrt{Var[\mathrm{x}]}/0.159$$

$$\epsilon \cong \sqrt{Var[\mathrm{x}]} \Big/ 0.159 \times L_{one\_round}$$

Riemannian Manifold Hamiltonian Monte Carlo (RMHMC)

RMHMC can semi-automatically tune the Leapfrog integration process

HMC $K(\mathbf{p}) \sim N(0, \mathbf{M})$ $\mathrm{M} \rightarrow \text{Identity matrix}$

⬇

RMHMC $K(\mathbf{p}) \sim N(0, \mathbf{G}(\boldsymbol{\theta}))$ $\mathbf{G}(\boldsymbol{\theta}) \rightarrow \text{Fisher information}$

Real Pig Data (Cleveland MA, et al., G3, 2012 2:429-435)

N	SNP	Mean	h2	Var(a)	$\hat{h}_g^2$	$\hat{H}^2$
3184	45385	37.99	0.62	3459.09	0.38	0.44

Da et al., PLoS One, 2014 9:e87666.

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Zd} + \mathbf{e}$$

Likelihood $\mathbf{y}|else \sim N(\mathbf{Xb} + \mathbf{Za} + \mathbf{Zd}, \mathbf{I}\sigma_e^2)$

Priors

$\mathbf{b}|\sigma_b^2 \sim N(0, 100^2)$ $\quad \sigma_e^2|\upsilon_e, S_e^2 \sim \chi_e^{-2}(\upsilon_e, S_e^2)$

$\mathbf{a}|\mathbf{G}, \sigma_a^2 \sim N(0, \mathbf{G}\sigma_a^2)$ $\quad \sigma_a^2|\upsilon_a, S_a^2 \sim \chi_a^{-2}(\upsilon_a, S_a^2)$

$\mathbf{d}|\mathbf{D}, \sigma_d^2 \sim N\left(0, \mathbf{D}\sigma_d^2\right)$ $\quad \sigma_d^2|\upsilon_d, S_d^2 \sim \chi_d^{-2}\left(\upsilon_d, S_d^2\right)$

Several setting

Fortran 90 (No use of special library (LAPACK, MKL,..,) within all McMC loops)

Random number generators

Uniform by Mersenne Twister: Piatek, R; Gamma: Dagpunar, J; Normal by Box Muller: Arakawa, A

Method	Distribution	ϵ	L
HMC	Normal	$\epsilon = \sqrt{Var[\mathrm{x}]} / 0.159 \times 20$	7
	χ^{-1}	$\epsilon = \sqrt{Var[\mathrm{v}]} / 0.112 \times 20$	
		$(L_{one\_round} = 20)$	
RMHMC		0.1	20

Method	
Effect $(\mathbf{b}, \mathbf{a}, \mathbf{d})$	Variance $(\sigma_a^2, \sigma_d^2, \sigma_e^2)$
Gibbs	Gibbs
Gibbs	HMC
Gibbs	RMHMC
HMC	HMC
HMC	RMHMC
RMHMC	RMHMC

5 MCMC chains with a different seed
Total sample=100,000, Burn-in= 10,000

Result: Estimates of genetic parameters

Method: Effects	Method: Variance	$\hat{\sigma}_a^2$	$\hat{\sigma}_d^2$	$\hat{\sigma}_e^2$	$\hat{h}^2$	$\hat{H}^2$
Gibbs	Gibbs	1301.8 (112.3)	159.2 (67.2)	2007.9 (85.8)	0.37 (0.02)	0.42 (0.03)
Gibbs	HMC	1304.4 (123.3)	175.4 (66.2)	1993.3 (91.8)	0.38 (0.03)	0.43 (0.03)
Gibbs	RMHMC	1302.9 (123.6)	175.0 (64.2)	1993.8 (91.1)	0.37 (0.03)	0.43 (0.03)
HMC	HMC	1308.0 (123.3)	173.5 (66.4)	1993.9 (92.1)	0.38 (0.03)	0.43 (0.03)
HMC	RMHMC	1307.4 (123.0)	171.0 (69.3)	1996.6 (93.7)	0.38 (0.03)	0.42 (0.03)
RMHMC	RMHMC	1307.0 (123.9)	163.8 (69.6)	2002.9 (94.8)	0.38 (0.03)	0.42 (0.03)

Result: Sampling Properties

Method		Effective Sample Size					Time (h)
Effects	Variance	$\hat{\sigma}_a^2$	$\hat{\sigma}_d^2$	$\hat{\sigma}_e^2$	$\hat{h}^2$	$\widehat{H}^2$	
Gibbs	Gibbs	1751.2	118.2	600.6	1749.9	856.7	3.53
Gibbs	HMC	2312.2	213.4	970.2	2247.1	1158.1	3.55
Gibbs	RMHMC	2131.5	208.6	994.7	2097.5	1180.0	3.55
HMC	HMC	2748.3	170.1	922.1	3008.9	1243.2	3.57
HMC	RMHMC	2596.3	147.4	797.4	2769.3	1093.4	3.56
RMHMC	RMHMC	2656.2	169.6	838.8	2782.5	1116.5	3.59

MacBook Pro (2.7 GHz Intel Core i7; 16 GB 2133 MHz LPDDR3)

Summary

HMC methods work well on the animal breeding data

$$\mathrm{ESS}_{GS} < \mathrm{ESS}_{RMHMC} = \mathrm{ESS}_{HMC}$$

$$\mathrm{Time}_{GS} = \mathrm{Time}_{RMHMC} = \mathrm{Time}_{HMC}$$

Remarkable property of the HMC

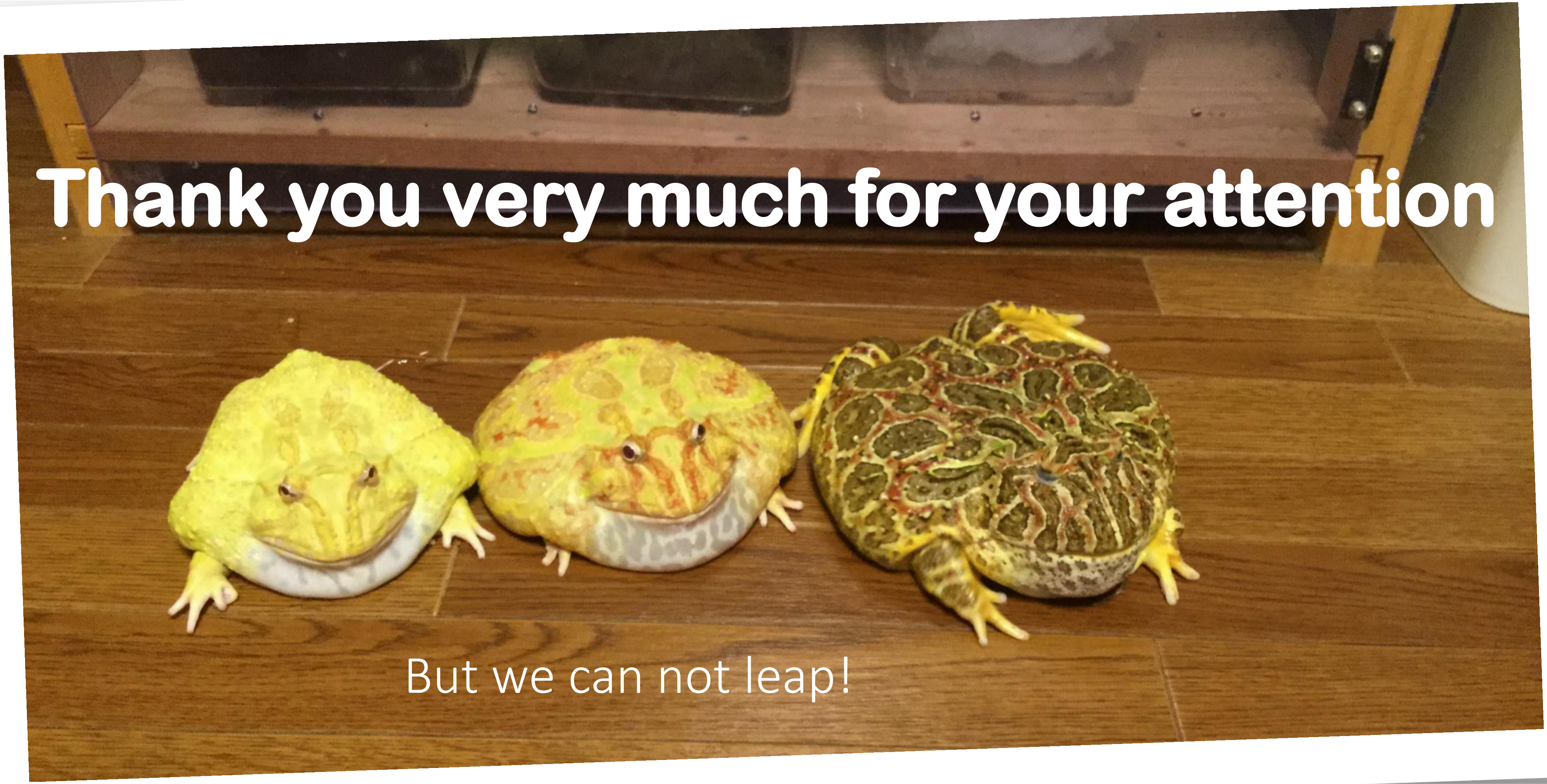
The HMC methods require no conjugate priors to a likelihood

≒ Metropolis-Hastings method

Acknowledgement

NARO

INRA
SCIENCE & IMPACT


Thank you very much for your attention
But we can not leap!