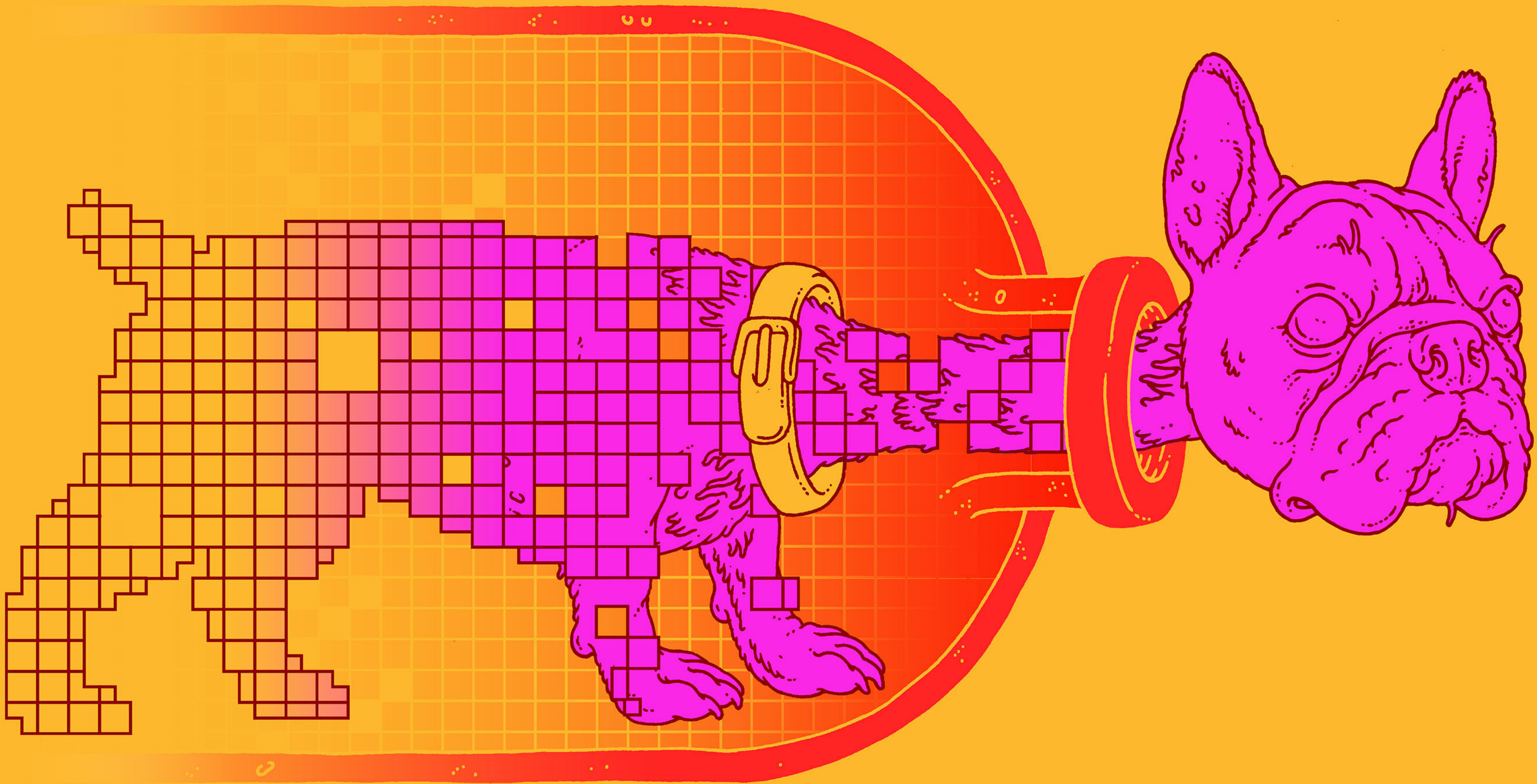


THE EARLY UNIVERSE INFORMATION BOTTLENECK

Jonathan Frazer (DESY)

string_data Workshop Munich 26 March 2018

OVERVIEW

How can we make predictions when the inflation model is not simple?

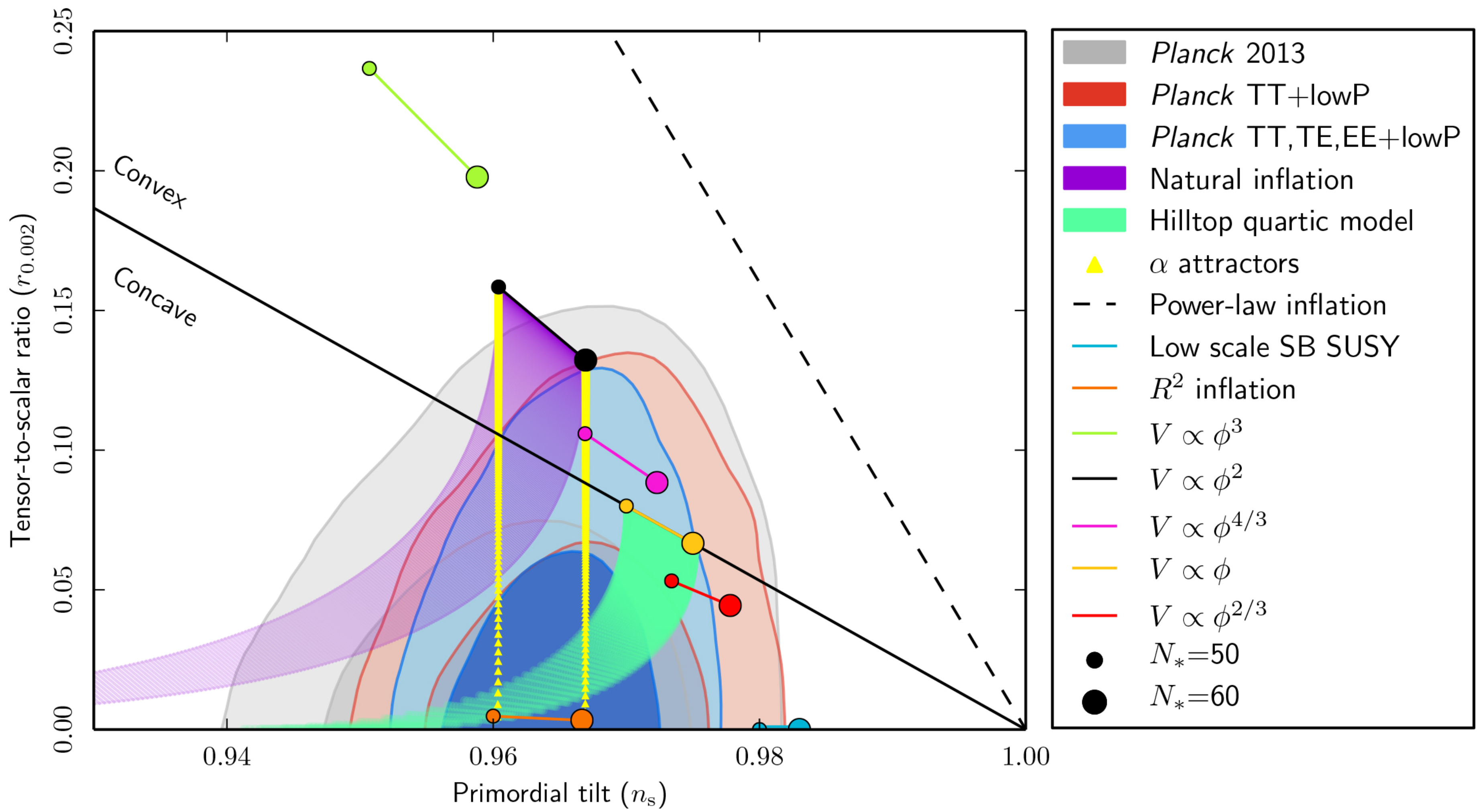
1. Computing observables from inflation models
2. Brief summary of publicly available codes
3. An information theoretic approach to making robust predictions
4. Minimal working example: single field axion monodromy

title page image by Eric Nyquist for Quanta Magazine

PREDICTIONS FROM SIMPLE MODELS

$$n_s(k) - 1 = -6\epsilon + 2\eta$$

$$r = 16\epsilon$$



$$P_t = 8\left(\frac{H}{2\pi}\right)^2 \quad r \equiv \frac{P_t}{P_\zeta} \quad \epsilon \equiv \frac{1}{2}\left(\frac{V'}{V}\right)^2 \quad \eta \equiv \frac{V''}{V}$$

Planck 2015 results. XX: arXiv:1502.02114

WHAT ABOUT WHEN THE MODEL IS NOT SO SIMPLE?

Single field models may require a numerical approach if

- Model has multiple parameters
- There are deviations from slow-roll when relevant scales exit the horizon

Models with more than than one "active" field (multifield) typically require numerics as dynamics are much richer
e.g.

- Sensitivity to initial conditions
- Super-horizon evolution of observables due to isocurvature
- Particle production
- Non-Gaussianity

Name	Authors	Language	Capabilities
FieldInf	C. Ringeval P. Brax C. van de Bruck A. C. Davis	Fortran	multifield power spectrum
ModeCode MultiModeCode	M. J. Mortonson H. V. Peiris R. Easther L. C. Price J. Frazer J. Xu	Fortran	multifield power spectrum sampling
PyFlation	Huston K. A. Malik	Python	single field second order
BINGO	D. K. Hazra, L. Sriramkumar J. Martin	Fortran	single field power spectrum bispectrum
mTransport PyTransport CppTransport	M. Dias J. Frazer D. Mulryne D. Seery J. Ronayne S. Butchers	Mathematica Python C++	multifield curved field spaces power spectrum bispectrum (sampling coming soon)

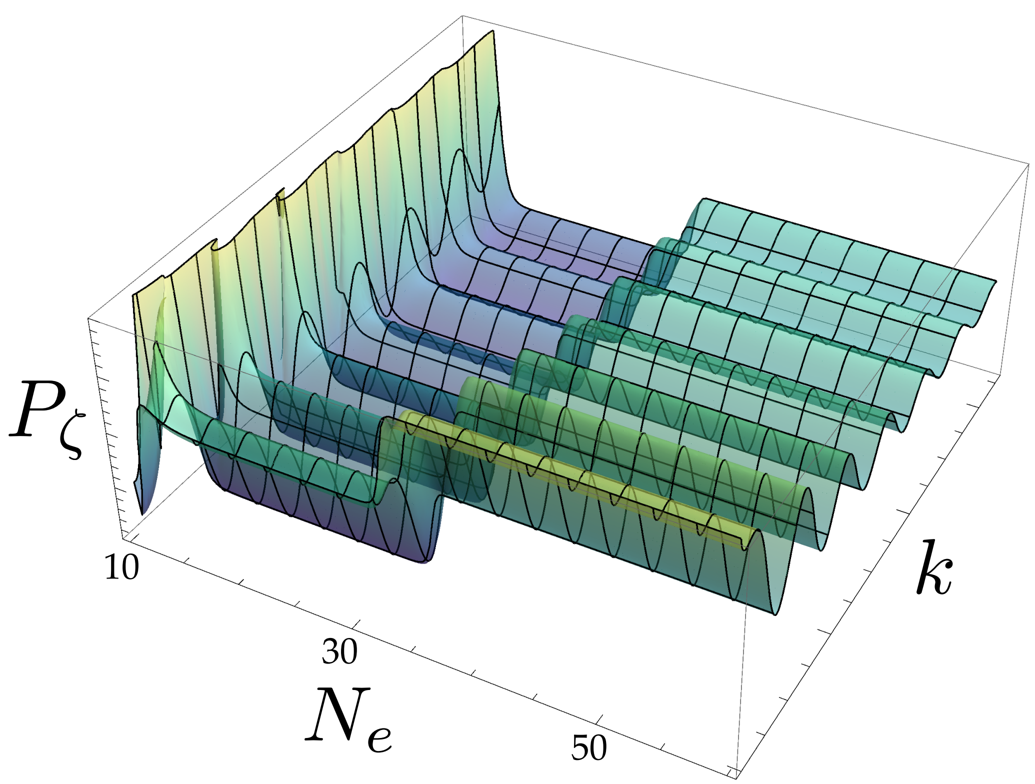
THE TRANSPORT METHOD

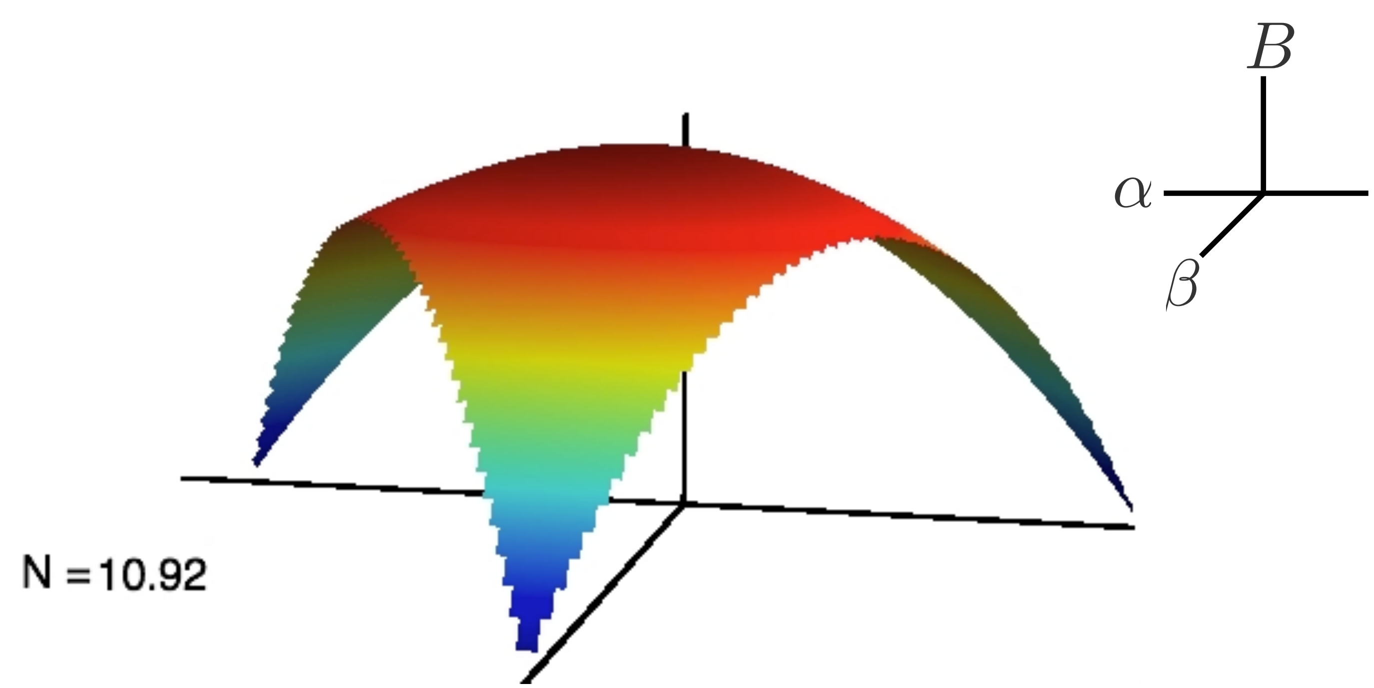
transportmethod.com

Solves ODEs for field and momenta correlation functions
Can handle models of the form

$$S = \frac{1}{2}\int \mathrm{d}^3x\mathrm{d}t\sqrt{-g}\left\{M_{\mathrm{p}}^2 R - G_{\alpha\beta}g^{\mu\nu}\partial_\mu\phi^\alpha\partial_\nu\phi^\beta - 2V\right\}$$

Example: $G^{\alpha\beta} = \begin{pmatrix} 1 & \Gamma & 0 \\ \Gamma & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ $\Gamma \equiv \frac{0.9}{\cosh\left((\phi_1^2 - 7)^2/0.06\right)}$ $V = \frac{1}{2}\sum_{\alpha=1}^{3} m_\alpha^2\phi_\alpha^2$





Bispectrum video created by Sean Butchers. Power spectrum: Dias, JF, Seery: arXiv: 1502.03125

TOWARDS ROBUST PREDICTIONS DESPITE INCOMPLETE KNOWLEDGE

Motivating Example: Random Matrix Theory

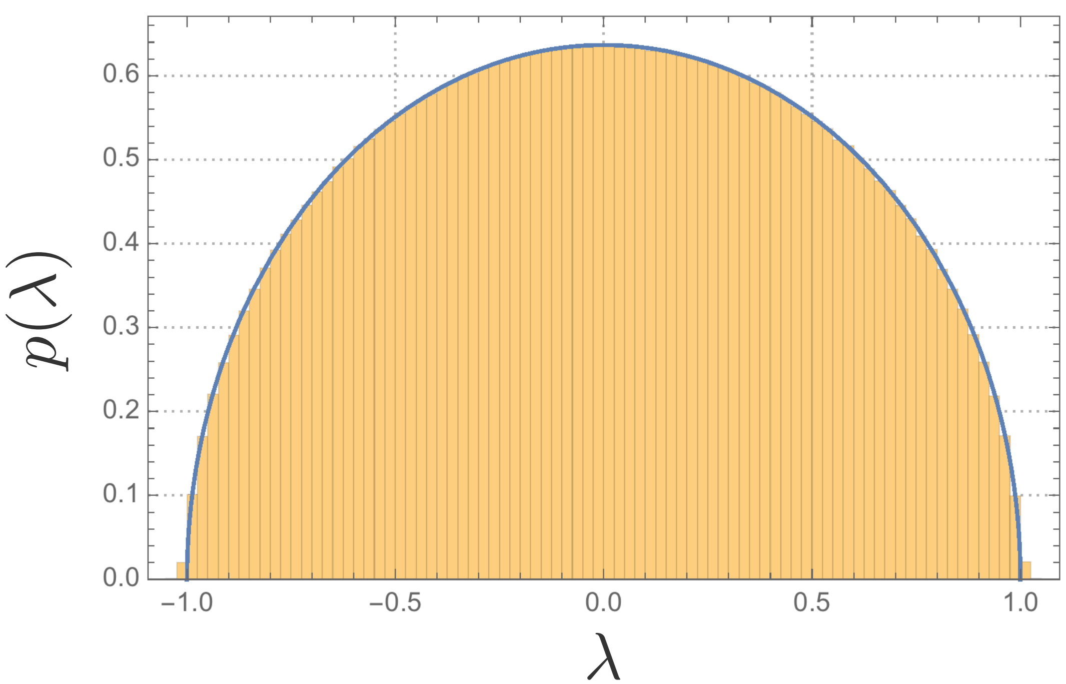
Random matrices first introduced to physics by Eugene Wigner
He modelled the nuclei of heavy atoms

Postulated that the spacings between the lines in the spectrum of a heavy atom nucleus should resemble the spacings between the eigenvalues of a random matrix, and should depend only on the symmetry class of the underlying evolution

Example:
Gaussian orthogonal ensemble

$$p(M)\mathrm{d}M = p(M')\mathrm{d}M'$$

$$M' = O^T M O$$



Can we find a general and systematic approach to testing for universality?

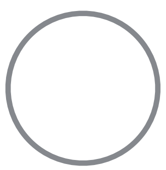
Mehta: Random Matrices

A NETWORK PERSPECTIVE ON INFLATION

$\alpha \; \beta \; N_e^*$ hyperparameters

ϕ_0 initial conditions

$b \;\; c$ parameters appearing in $\mathcal{L}$

 stochastic variable

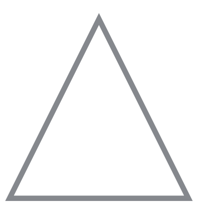
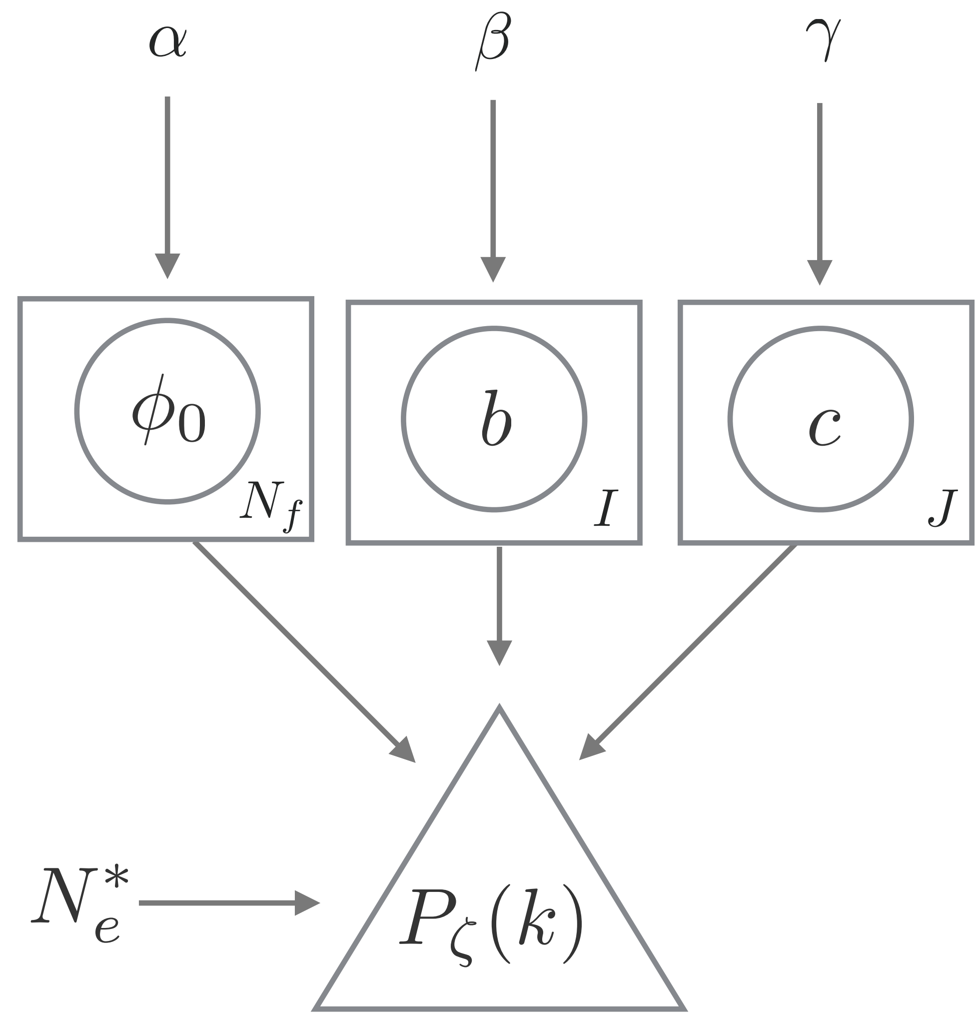
 deterministic variable

Plate: repeat n times



$$p(P_\zeta, \phi_0, b, c) = p(P_\zeta | \phi_0, b, c) p(\phi_0) p(b) p(c)$$

More generally $$p(X_1, \ldots, X_n) = \prod_{i=1}^{n} p(X_i | X_{\mathrm{pa(i)}})$$

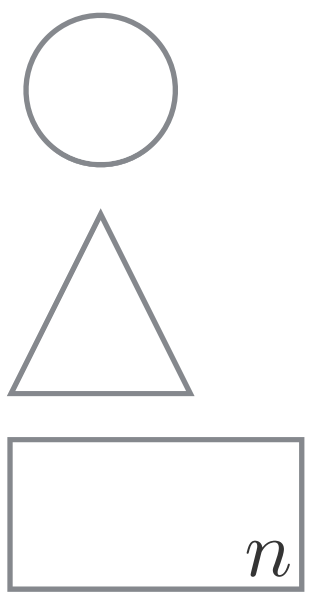
parent nodes

A NETWORK PERSPECTIVE ON INFLATION

$\alpha \ \beta \ N_e^*$ hyperparameters

ϕ_0 initial conditions

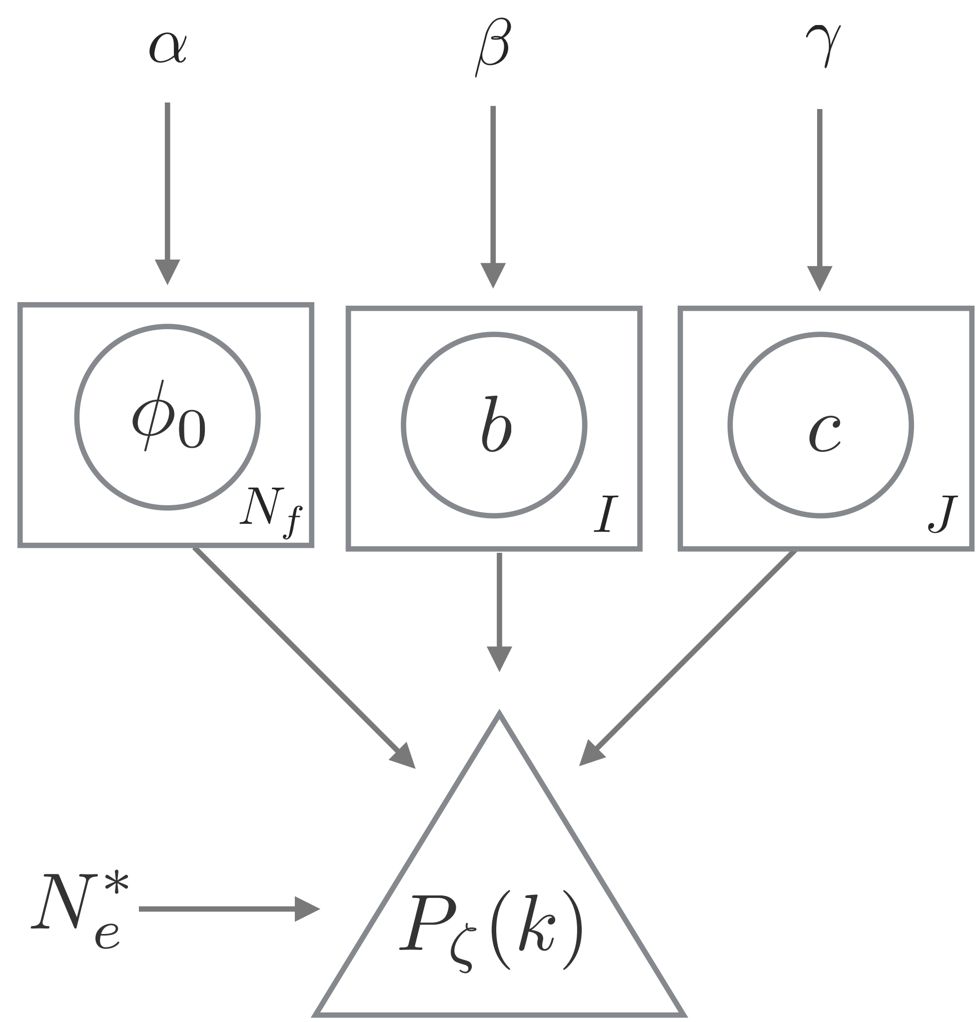
$b \ c$ parameters appearing in $\mathcal{L}$



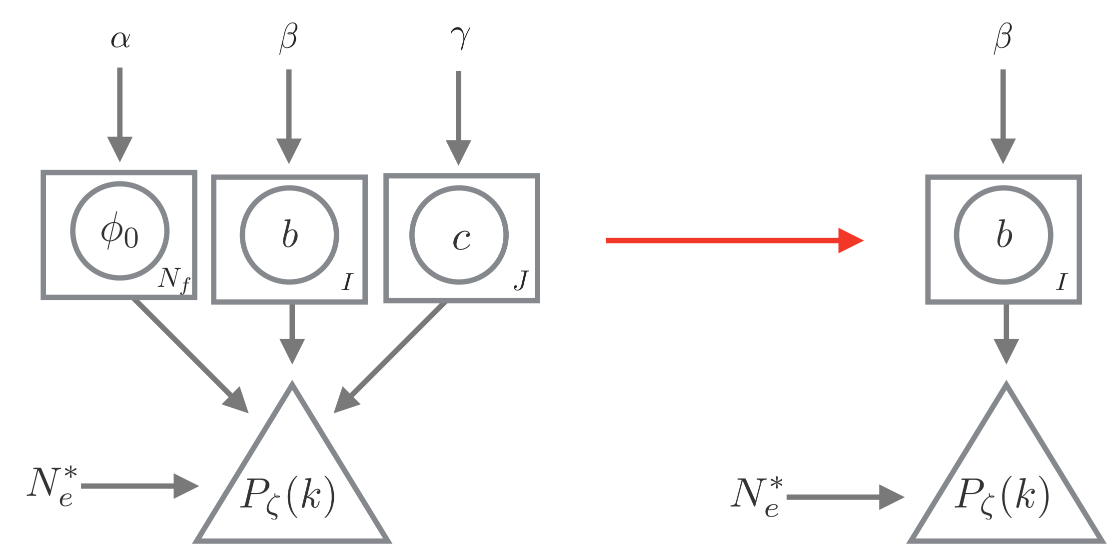
stochastic variable

deterministic variable

Plate: repeat n times



The task of making robust predictions is greatly simplified if we can show that some nodes can be safely “integrated out”

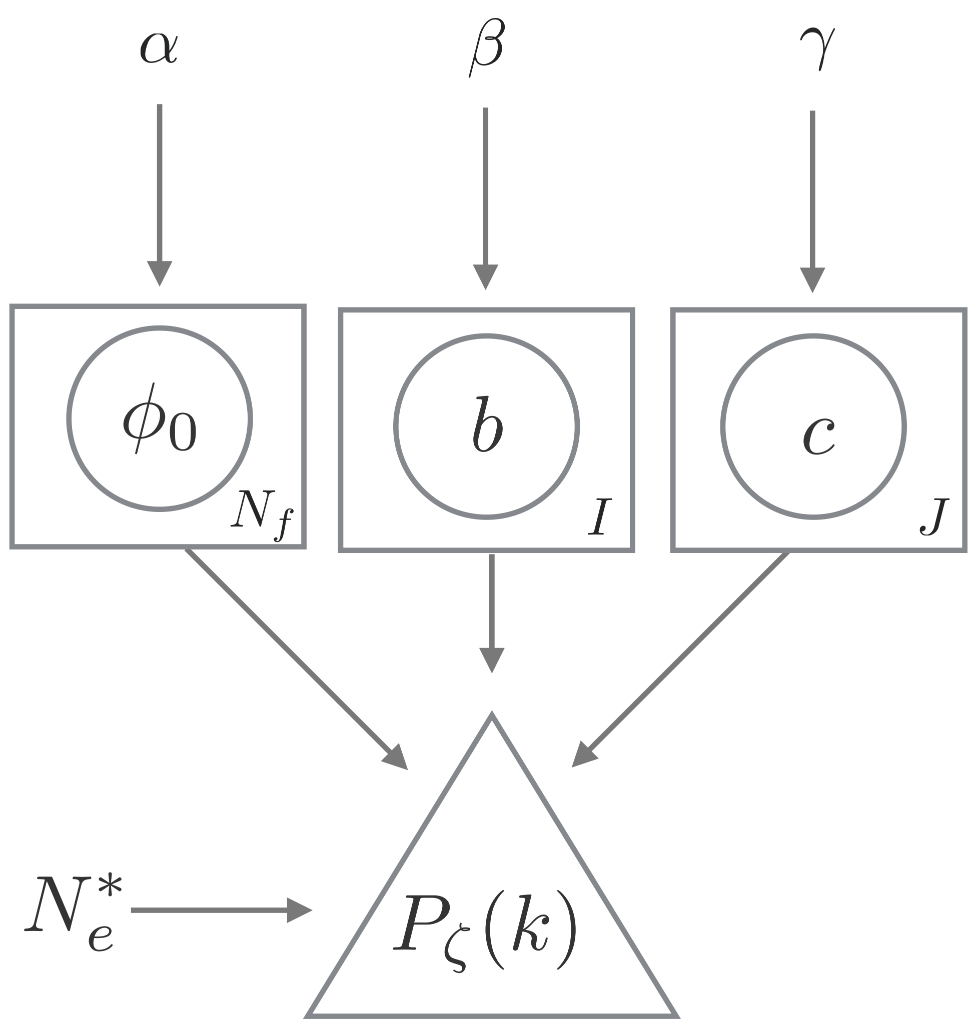


Objective: Asses the sensitivity of observables to choice of priors for model parameters (and hopefully identify hierarchies between the dependencies of the graph)

Steps:

1. Identify relevant scales (class of models)

   Model dependent but often one can obtain order of magnitude estimates for model parameters.



Objective: Asses the sensitivity of observables to choice of priors for model parameters (and hopefully identify hierarchies between the dependencies of the graph)

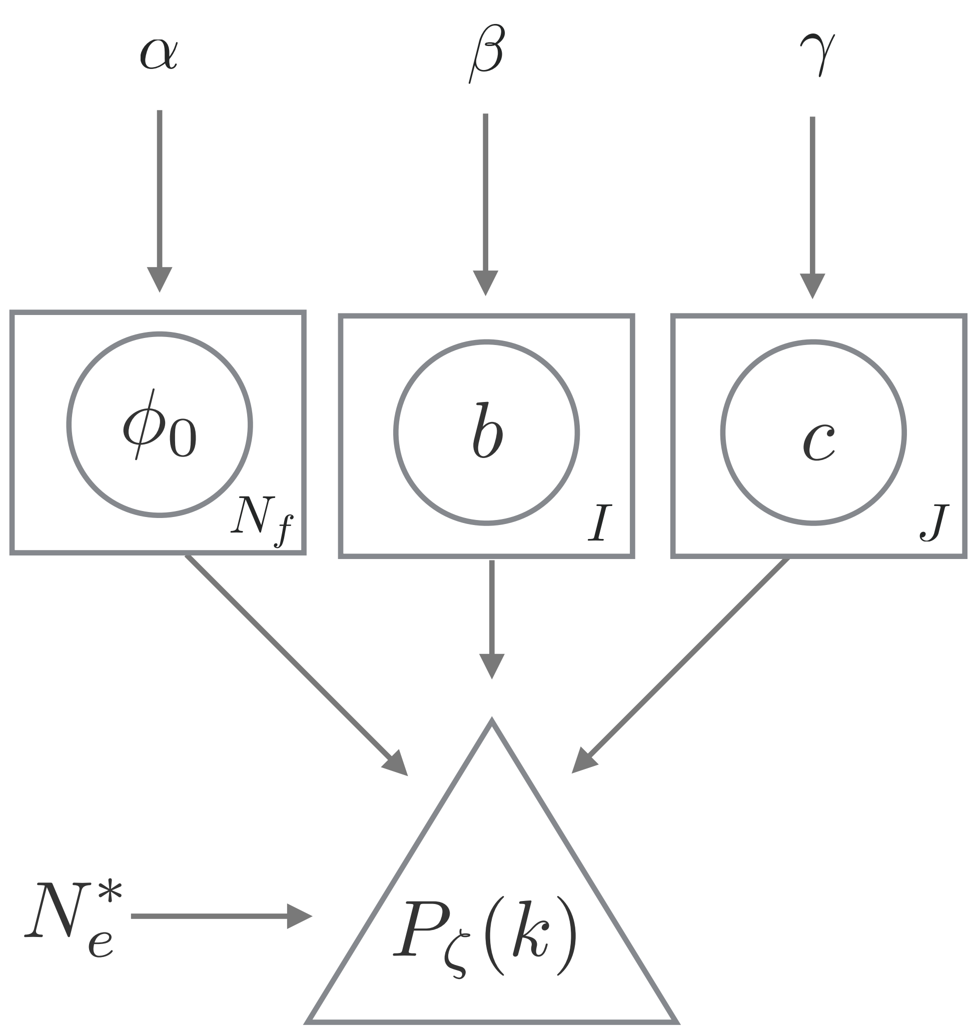
Steps:

1. Identify relevant scales (class of models)

   Model dependent but often one can obtain order of magnitude estimates for model parameters.

2. Learn the mapping from parameters to observables

   Use publicly available code to compute observables for large sample of model parameters.
   Use basic machine learning methods to learn this mapping



Objective: Asses the sensitivity of observables to choice of priors for model parameters (and hopefully identify hierarchies between the dependencies of the graph)

Steps:

1. Identify relevant scales (class of models)

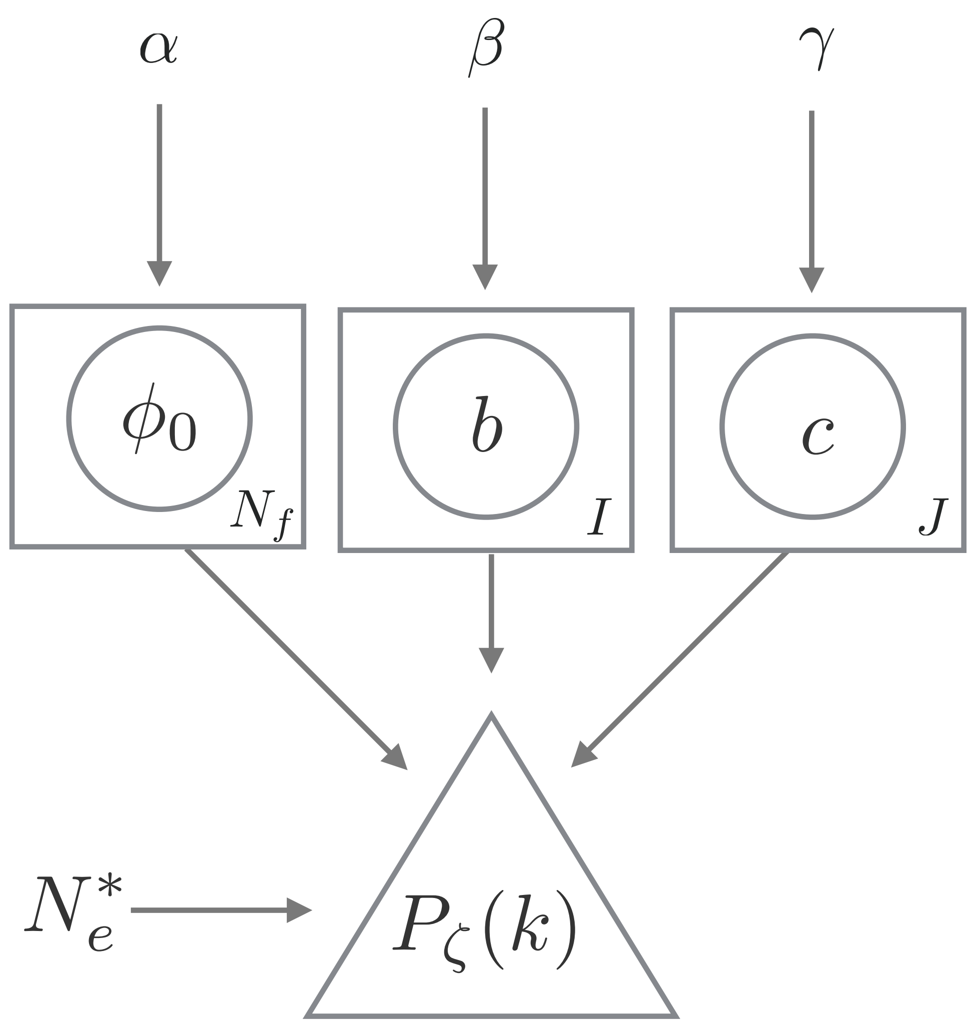
   Model dependent but often one can obtain order of magnitude estimates for model parameters.

2. Learn the mapping from parameters to observables

   Use publicly available code to compute observables for large sample of model parameters.
   Use basic machine learning methods to learn this mapping

3. Study how predictions depend on prior choice

   One approach is to use information theory



MUTUAL INFORMATION

Measures the average information x conveys about y.
Symmetric, non-negative, reparameterisation independent.

$$I(X;Y) \equiv \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

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

(marginal) entropy — conditional entropy

Shannon information

$$H(X) \equiv \sum_{x \in X} p(x) \log \frac{1}{p(x)}$$

$$H(X|Y) \equiv \sum_{y \in Y} p(y) H(X|Y=y)$$

THE DATA PROCESSING THEOREM

Consider an ensemble WDR where $w \rightarrow d \rightarrow r$ is a Markov chain

$$I(W;R) \leq I(W;D)$$

"data processing can only destroy information"

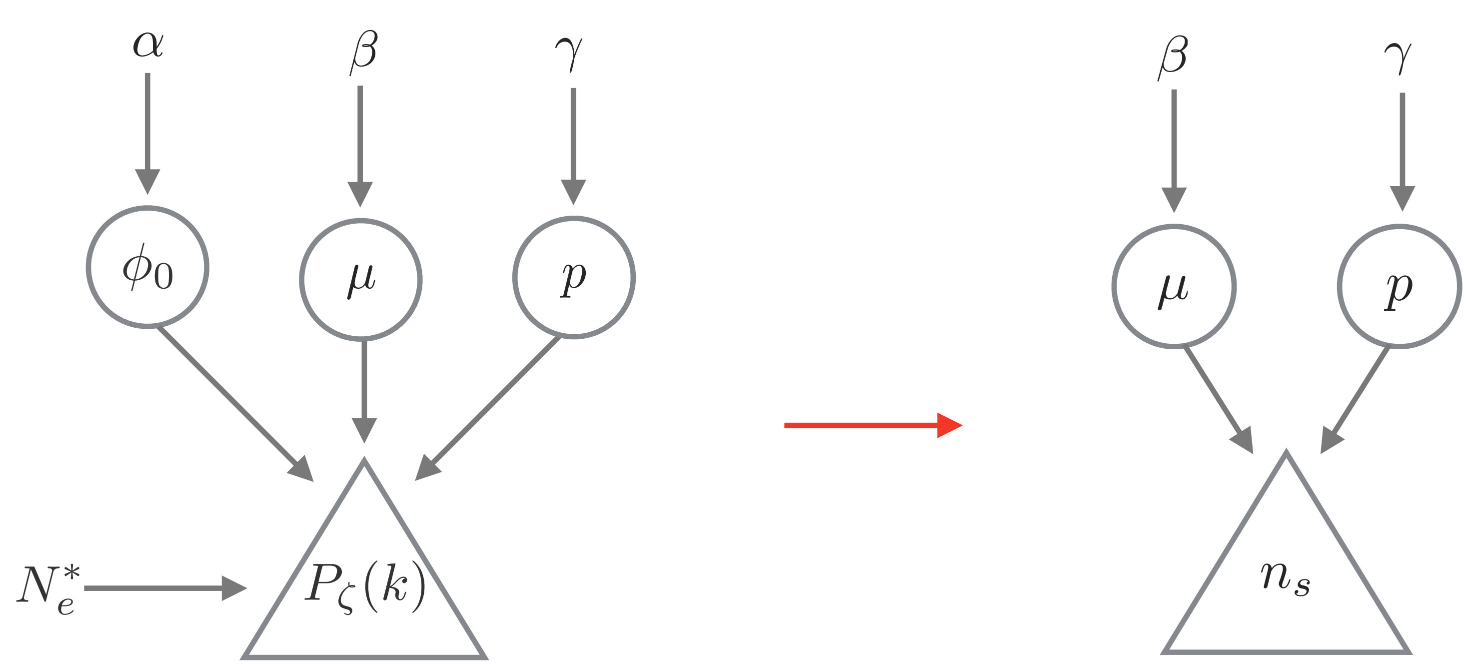
$$I(X;Y) \equiv \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

MINIMAL WORKING EXAMPLE: AXION MONODROMY

$$\mathcal{L} = -\frac{1}{2}(\partial\phi)^2 - \Lambda^4 \left[\left(1 + \left(\frac{\phi}{\mu} \right)^2 \right)^{\frac{p}{2}} - 1 \right]$$

Simplifications:

- Only consider one observable
- Include controlled backreaction, and assume sufficient inflation is always achieved
- Slow-roll initial conditions (but allow subsequent dynamics to be non slow-roll)
- Assume instantaneous reheating

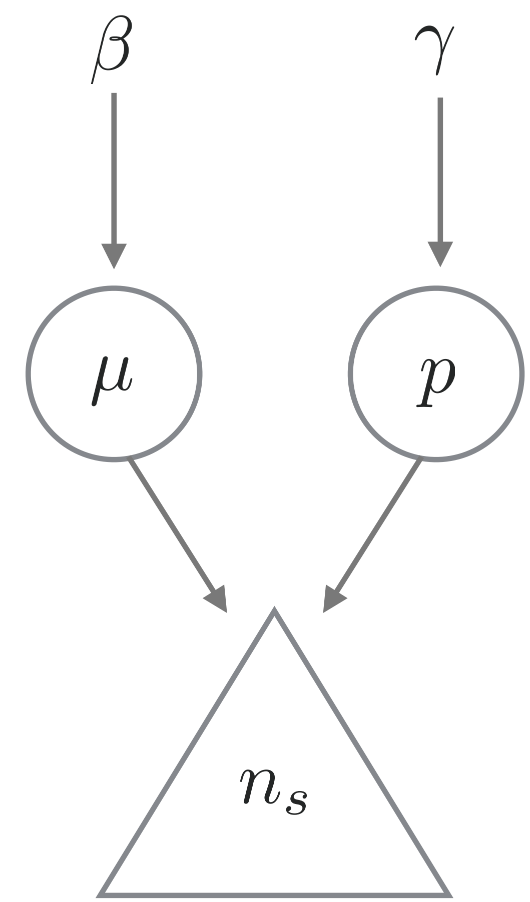


See Baumann and McAllister “Inflation and String Theory” for a review

MINIMAL WORKING EXAMPLE: AXION MONODROMY

Steps:

1. Identify relevant scales (class of models)
2. Learn the mapping from parameters to observables
3. Study how predictions change according to prior choice



μ Current efforts relating to the weak gravity conjecture may lead to upper bound. In the case of NS5-brane construction

Expect $\mu \in [0.1, 1]$

$$\mathcal{L} = -\frac{1}{2}(\partial\phi)^2 - \Lambda^4 \left[\left(1 + \left(\frac{\phi}{\mu} \right)^2 \right)^{\frac{p}{2}} - 1 \right]$$

p Depends on the scenarios of interest and “flattening mechanisms”. Known examples range from 2/5 to 2. Here we choose

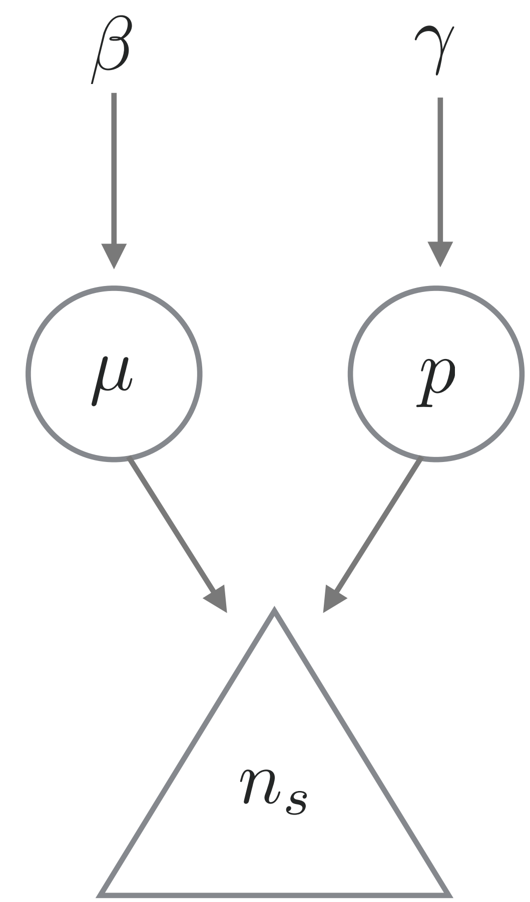
$$p \in [0.1, 2]$$

See Baumann and McAllister “Inflation and String Theory” for a review

MINIMAL WORKING EXAMPLE: AXION MONODROMY

Steps:

1. Identify relevant scales (class of models)
2. Learn the mapping from parameters to observables
3. Study how predictions change according to prior choice



Use numerical methods developed in previous work to generate a large sample assuming

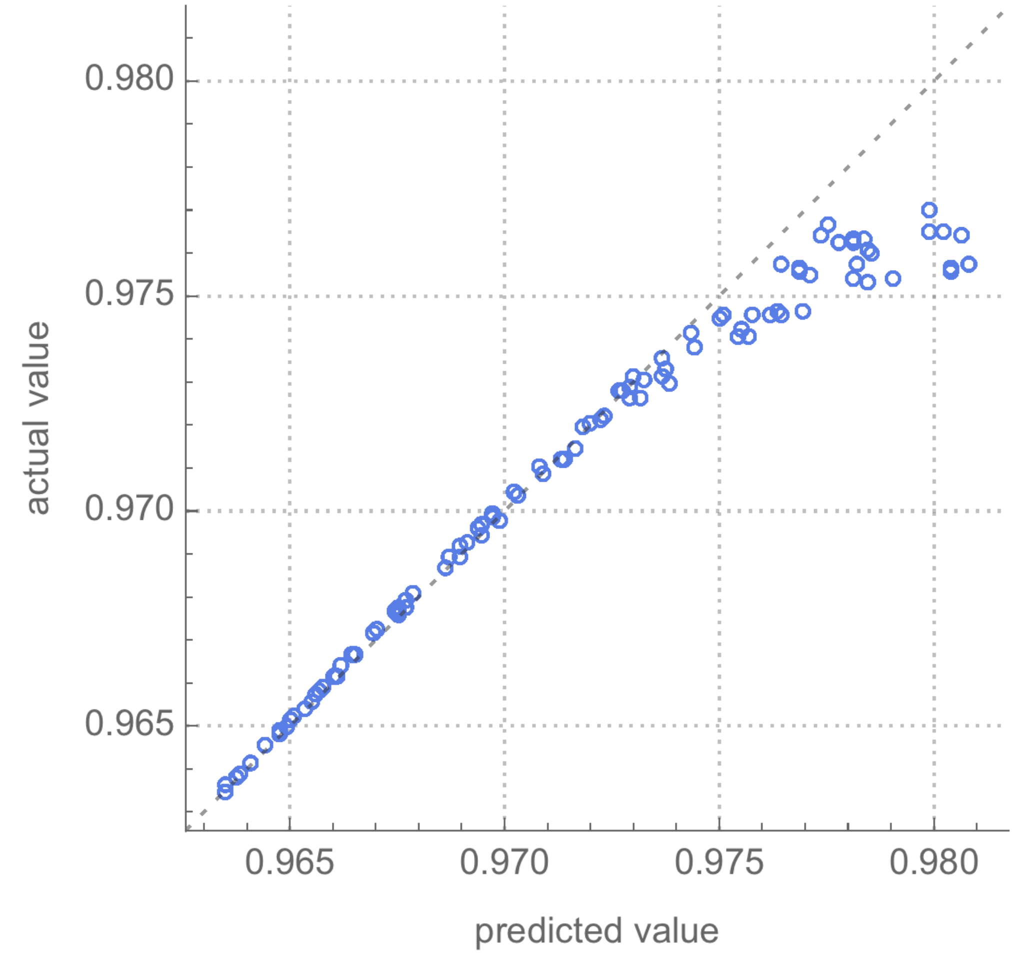
$$\mu \sim \mathcal{U}(0.1, 1) \quad p \sim \mathcal{U}(0.1, 2)$$

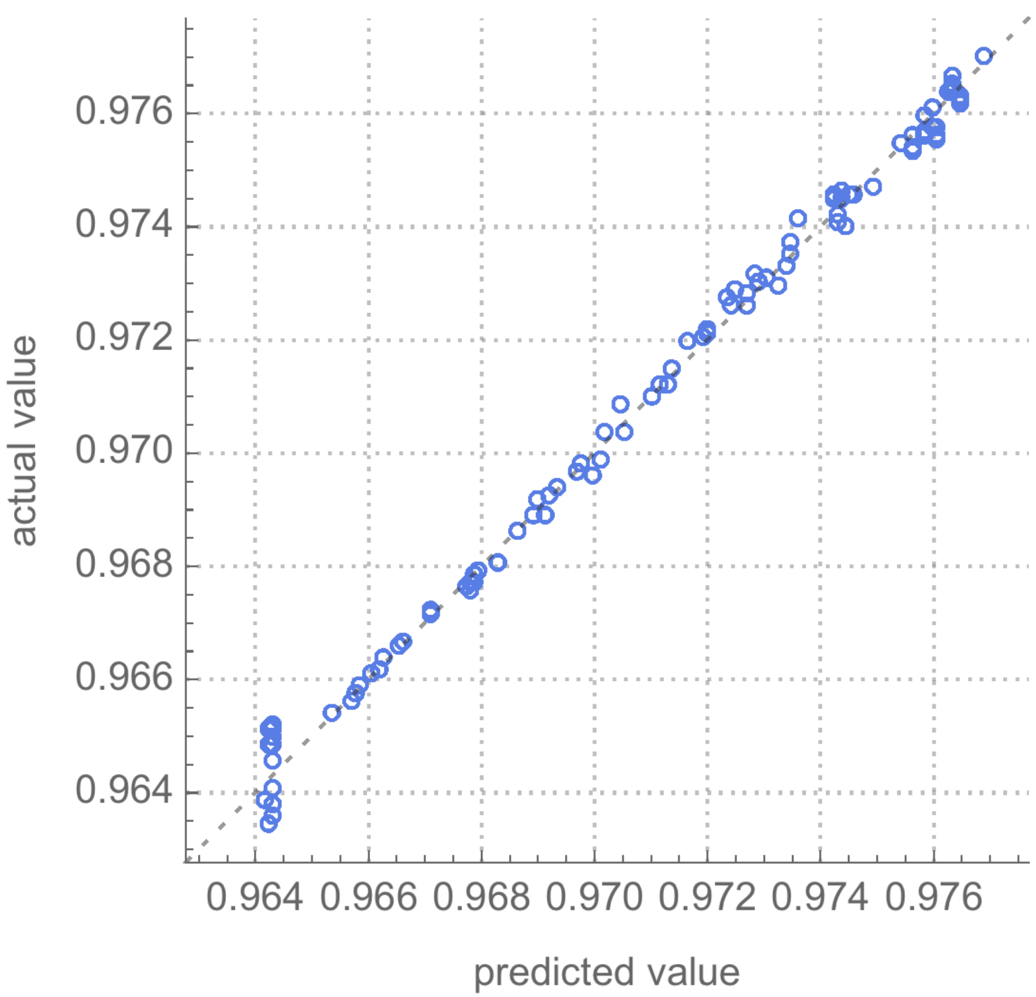
- Take a random draw of μ and p
- Solve background equations of motion to get total number of e-folds
- Solve equations of motion for the perturbations and compute n_s at pivot scale
- Repeat many times
- Use appropriate regression method to get $n_s(\mu, p)$

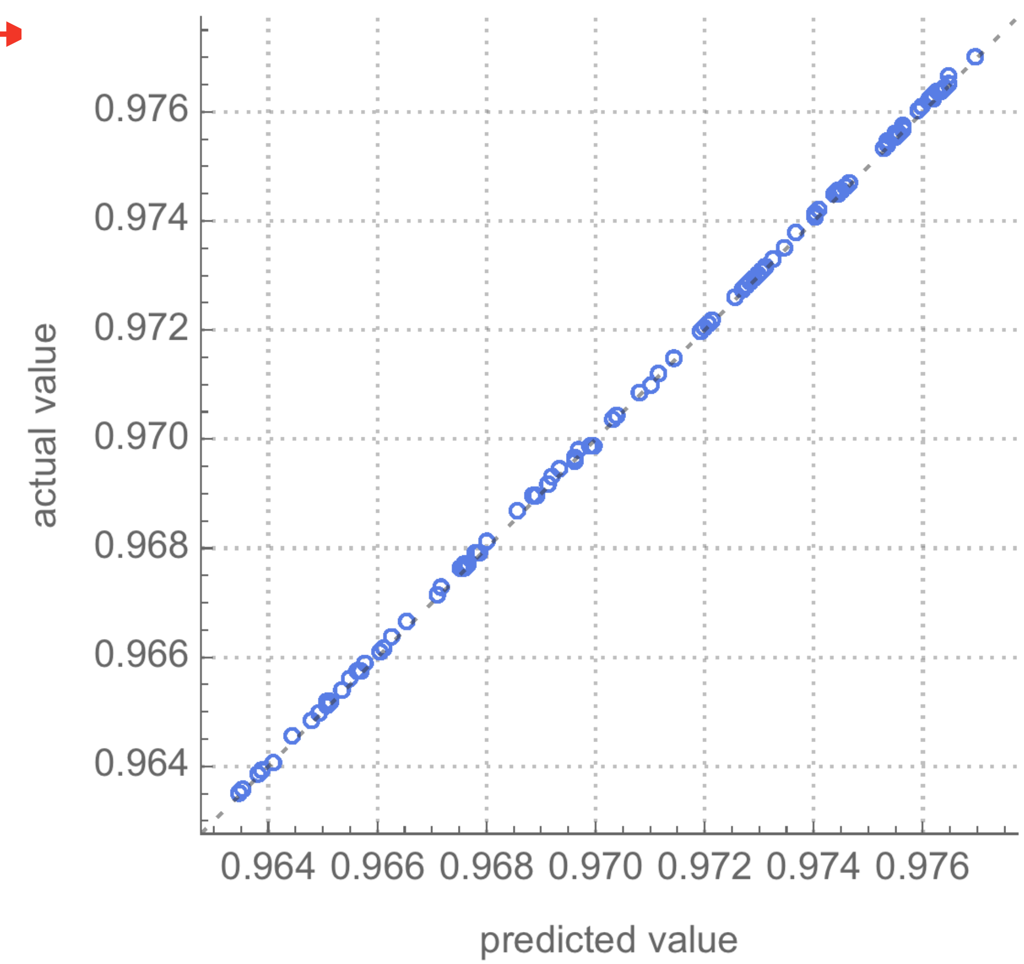
$$\mathcal{L} = -\frac{1}{2}(\partial\phi)^2 - \Lambda^4 \left[\left(1 + \left(\frac{\phi}{\mu} \right)^2 \right)^{\frac{p}{2}} - 1 \right]$$

Considered three regression methods:

- Neural Network
- Gaussian Process
- Gradient Boosted Trees



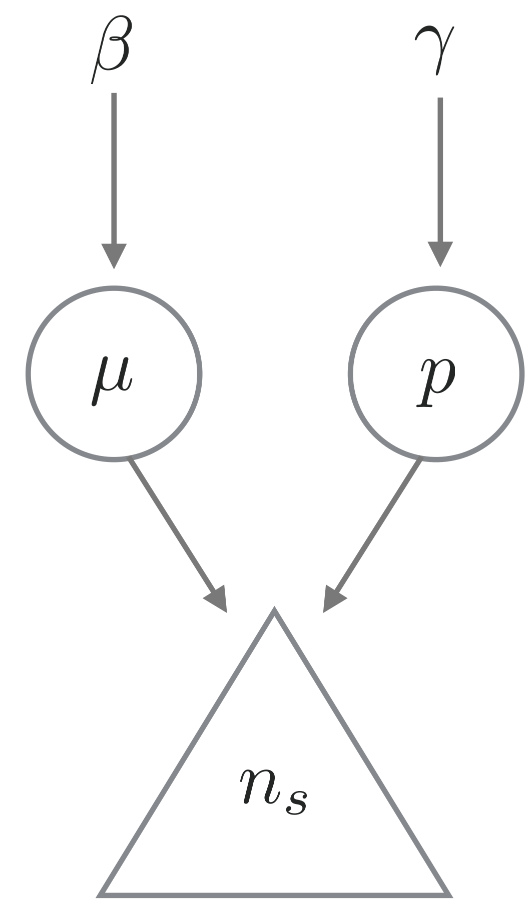




MINIMAL WORKING EXAMPLE: AXION MONODROMY

Steps:

1. Identify relevant scales (class of models)
2. Learn the mapping from parameters to observables
3. Study how predictions change according to prior choice



Consider a range of priors and compute the mutual information

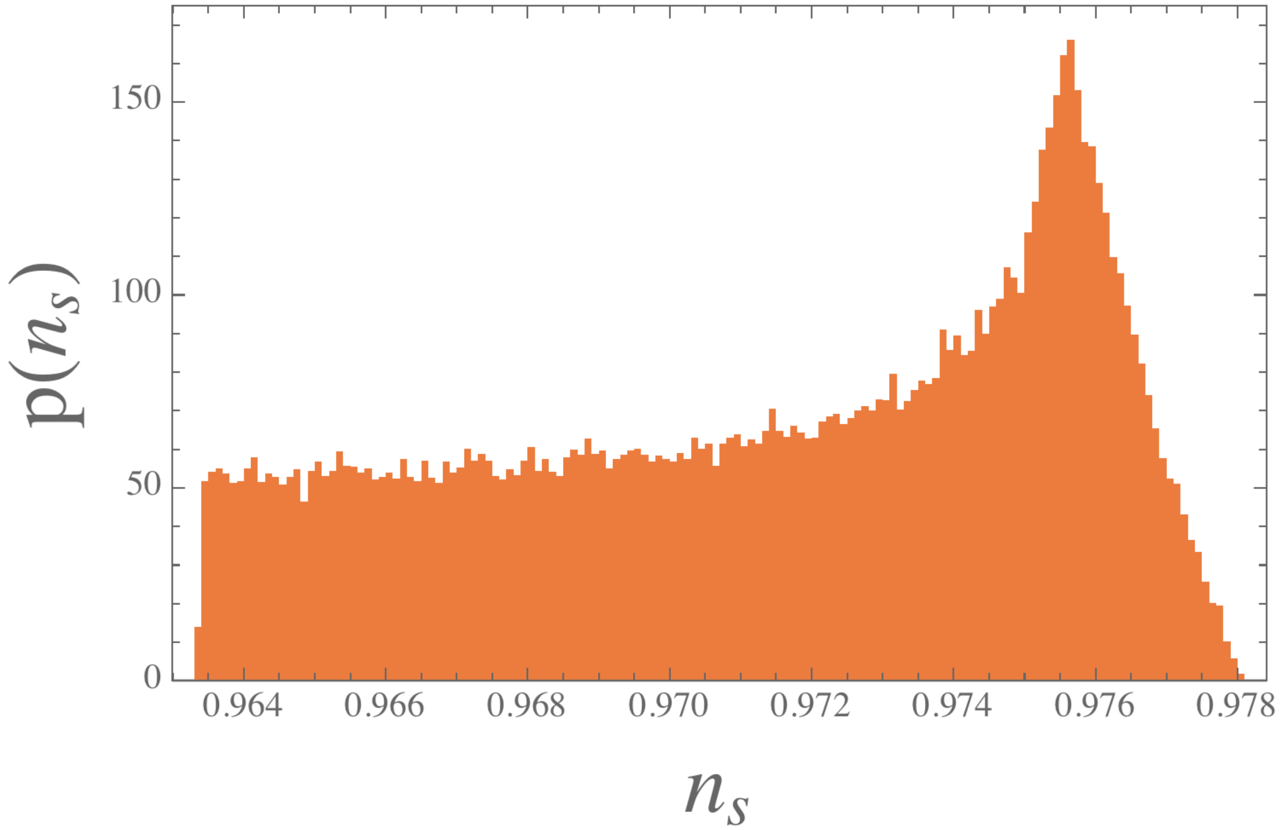
$$I(n_s;\mu) \quad I(n_s;p)$$

$$I(n_s;\mu) = \sum_{n_s}\sum_{\mu} p(n_s,\mu)\log\left[\frac{p(n_s,\mu)}{p(n_s)p(\mu)}\right]$$

$$= H(\mu) - H(\mu|n_s)$$

MINIMAL WORKING EXAMPLE: AXION MONODROMY

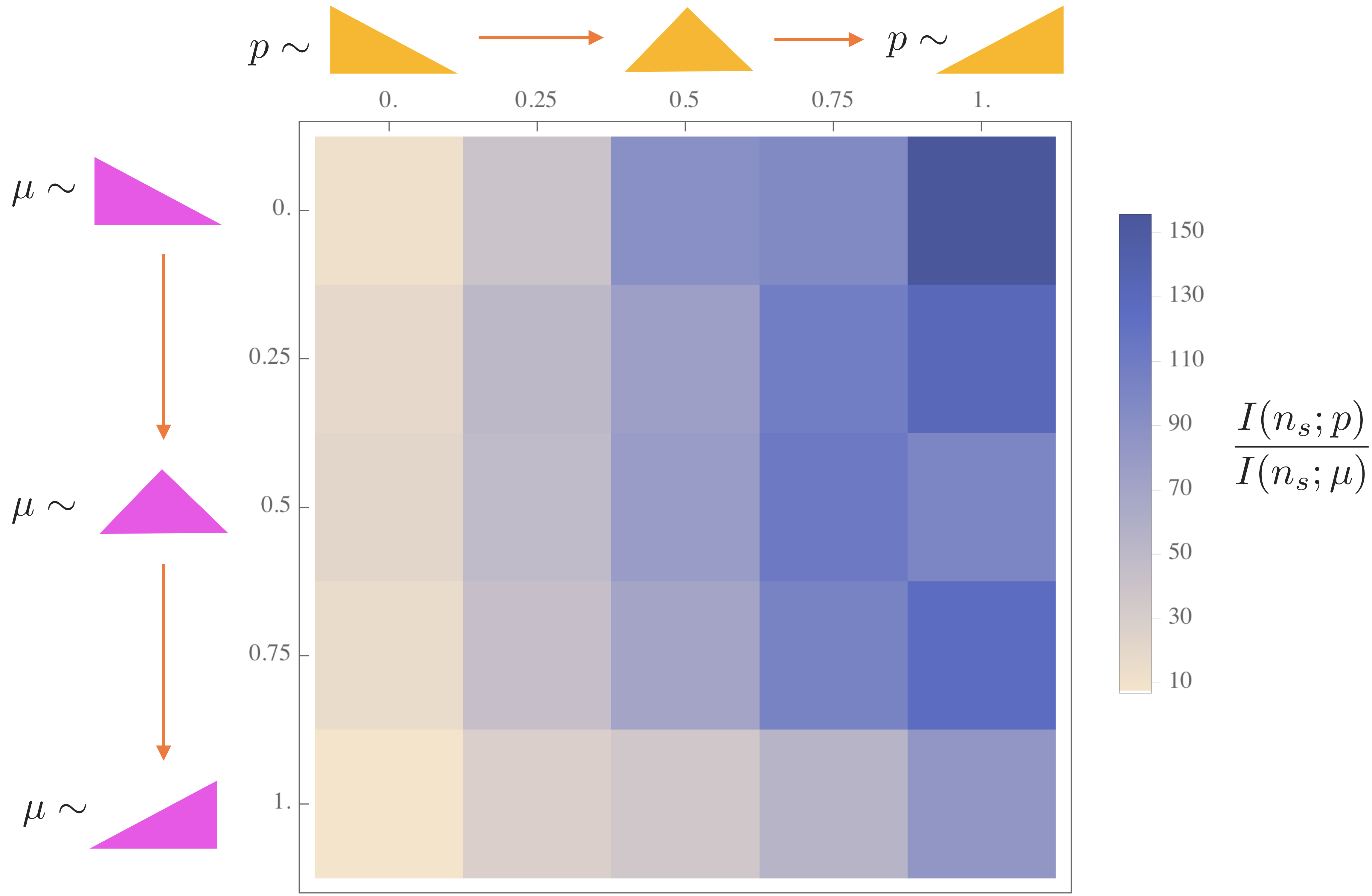
$\mu \sim$  $p \sim$ 



$$\frac{I(n_s; p)}{I(n_s; \mu)} \approx 17$$

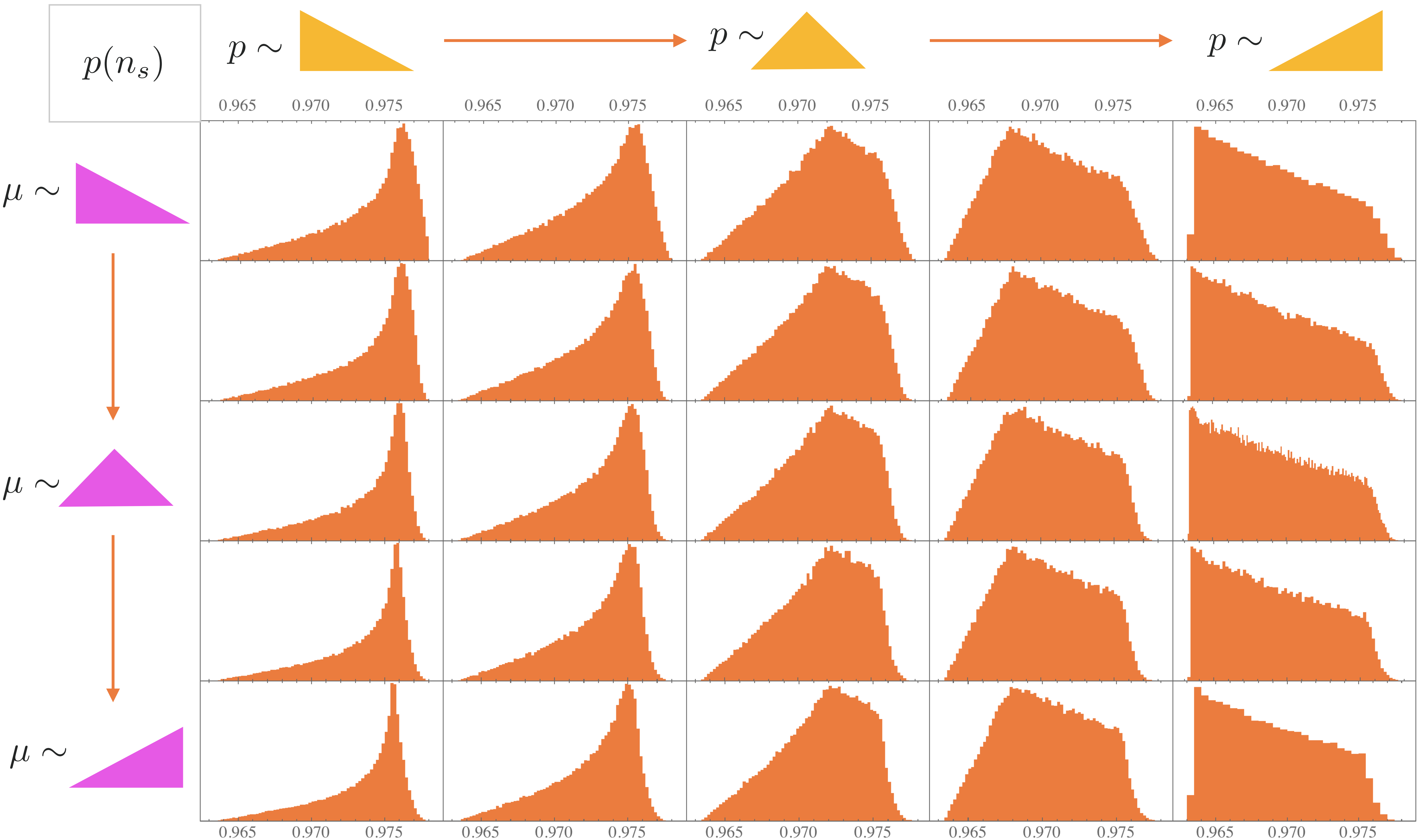
The spectral index contains significantly more information about p than μ

MINIMAL WORKING EXAMPLE: AXION MONODROMY



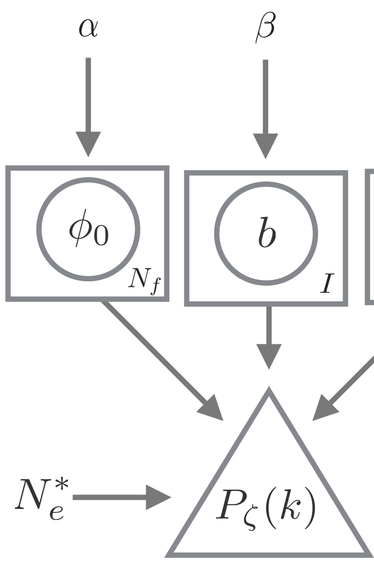
The spectral index contains significantly more information about p than μ

MINIMAL WORKING EXAMPLE: AXION MONODROMY

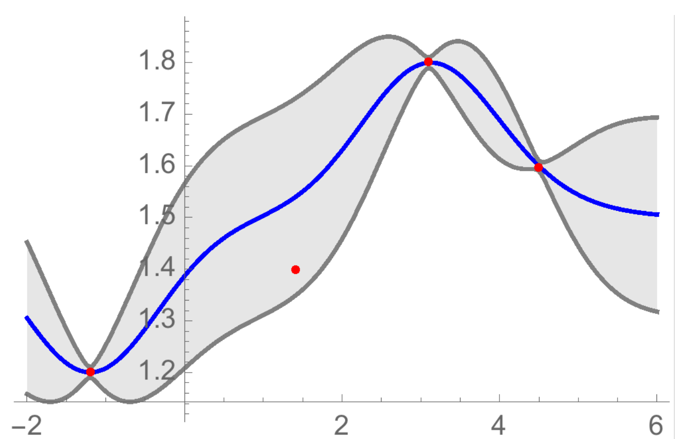


The spectral index contains significantly more information about p than μ

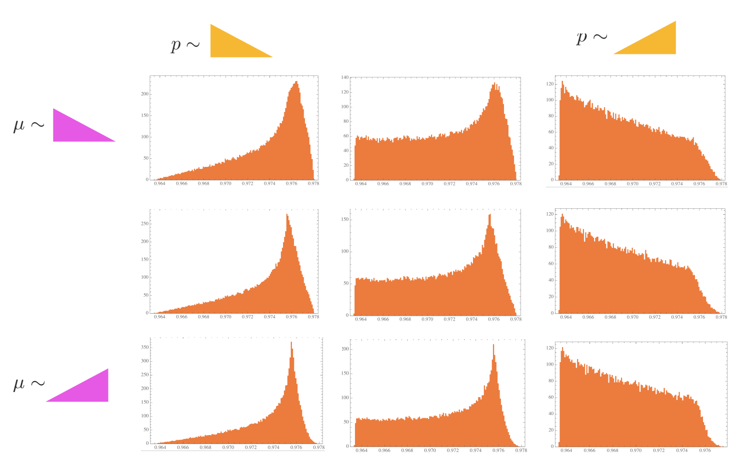
CONCLUSIONS



Bayesian networks are a highly flexible framework well suited for model building in string theory



Regression methods currently being developed by the machine learning community can radically reduce the computational cost of studying inflation models



The mapping from model parameters to observables results in an information bottleneck. This may enable robust predictions despite incomplete knowledge of the underlying theory