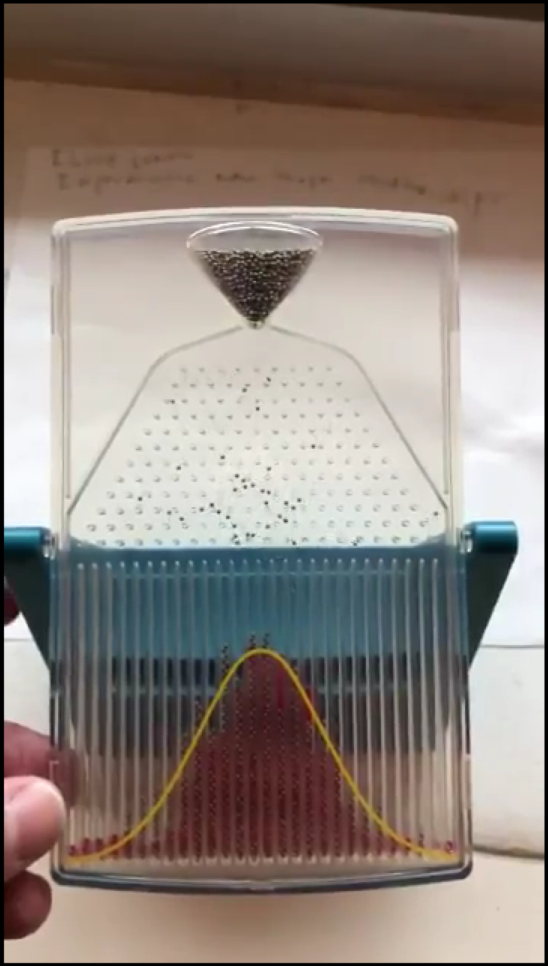
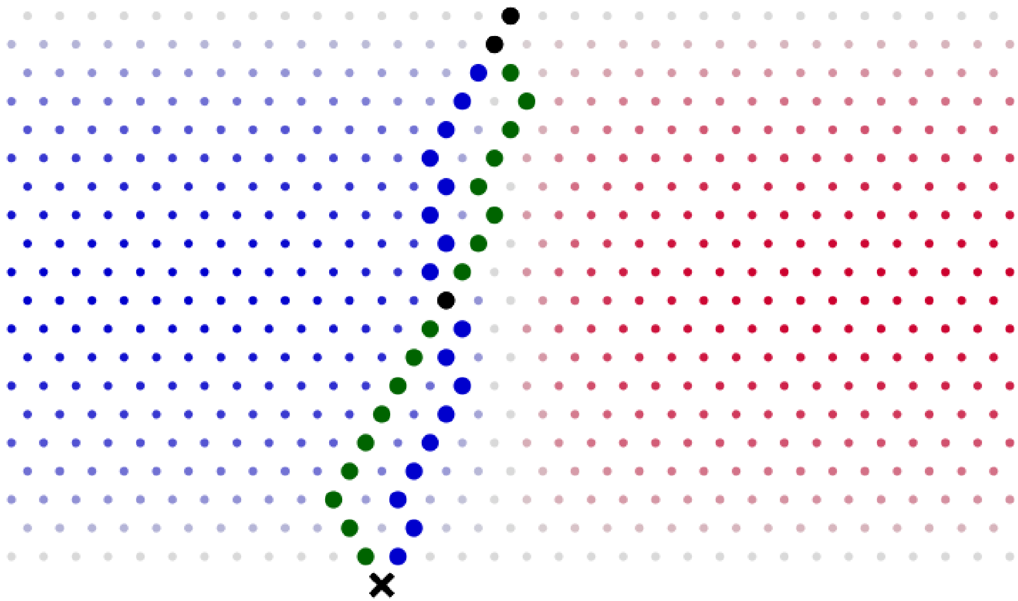
Likelihood-free inference in Physical Sciences

Artificial Intelligence and Physicss

Institut Pascal
March 22, 2019

Gilles Louppe
@glouppe





The probability of ending in bin x corresponds to the total probability of all the paths z from start to x.

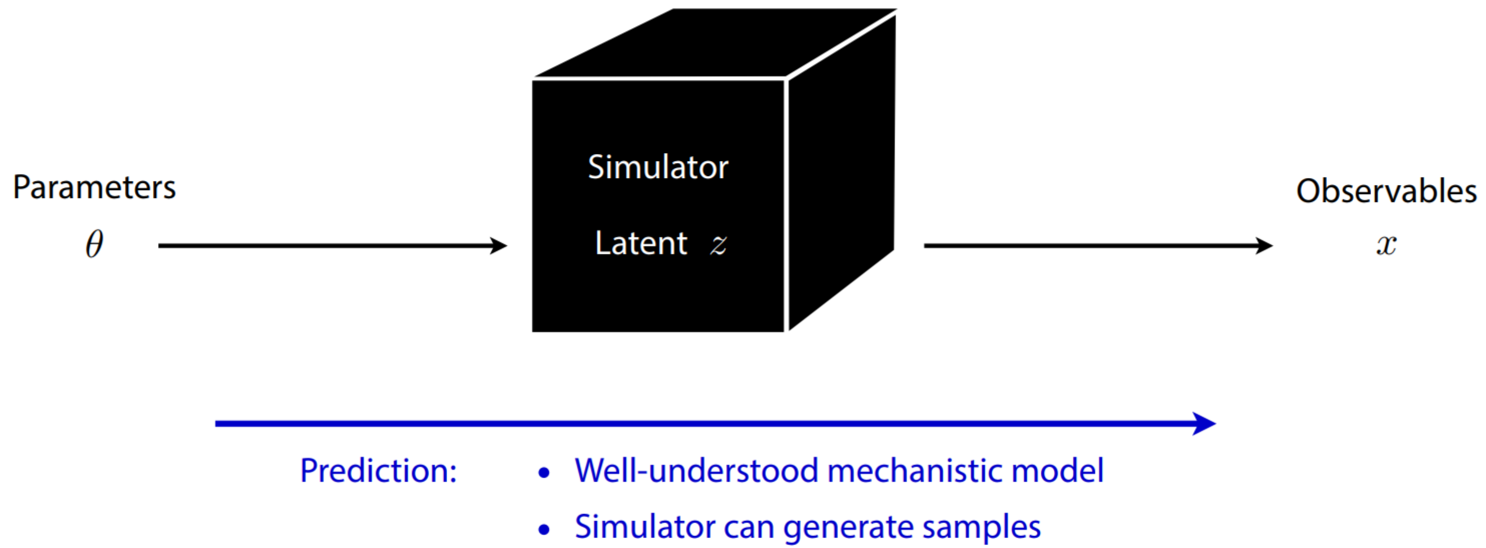
$$p(x|\theta) = \int p(x, z|\theta)dz = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

But what if we shift or remove some of the pins?

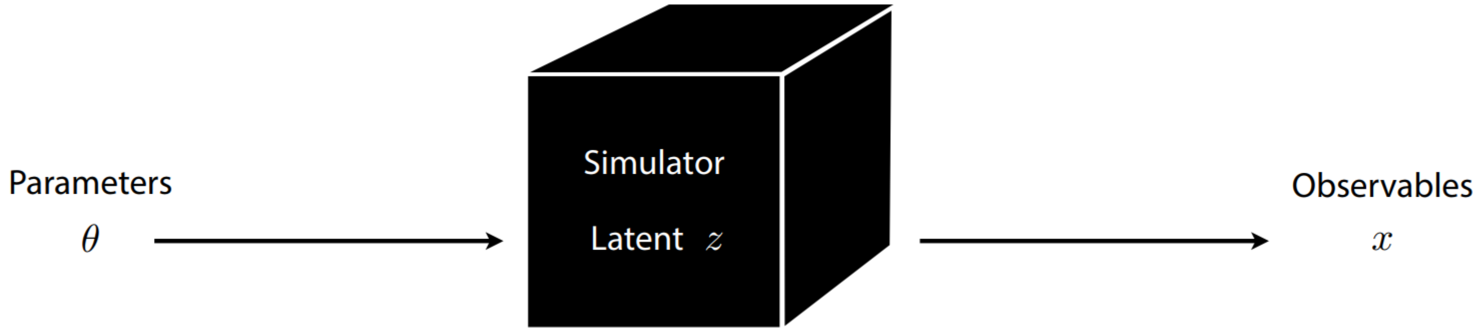
The Galton board is a metaphore of simulation-based science:

Galton board device	$\rightarrow$	Computer simulation
Parameters θ	$\rightarrow$	Model parameters θ
Buckets x	$\rightarrow$	Observables x
Random paths z	$\rightarrow$	Latent variables z (stochastic execution traces through simulator)

Inference in this context requires likelihood-free algorithms.



Credits: Johann Brehmer.



Prediction:

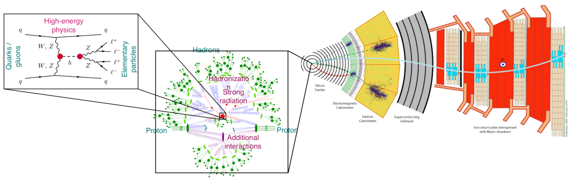
- Well-understood mechanistic model
- Simulator can generate samples

Inference:

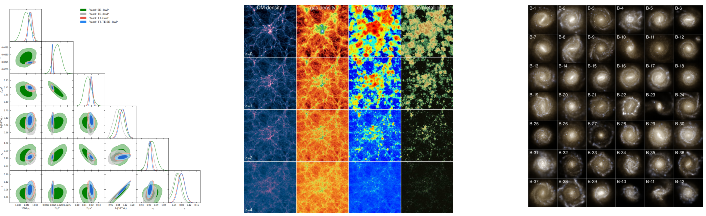
- Likelihood function $p(x|\theta)$ is intractable
- Inference based on estimator $\hat{p}(x|\theta)$

Credits: Johann Brehmer.

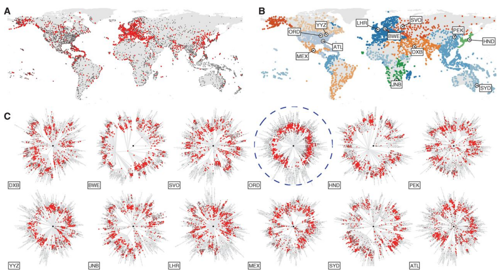
Applications



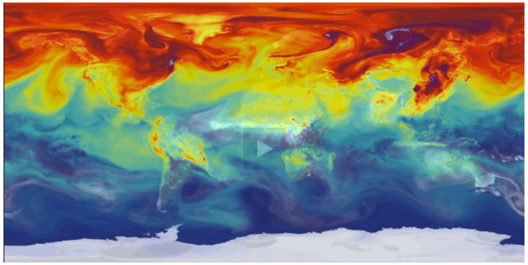
Particle physics



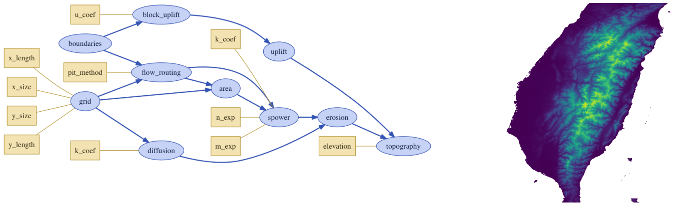
Cosmology



Epidemiology



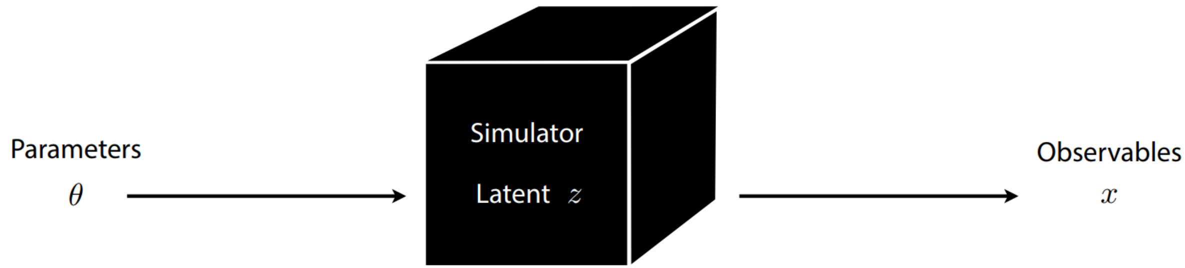
Climatology



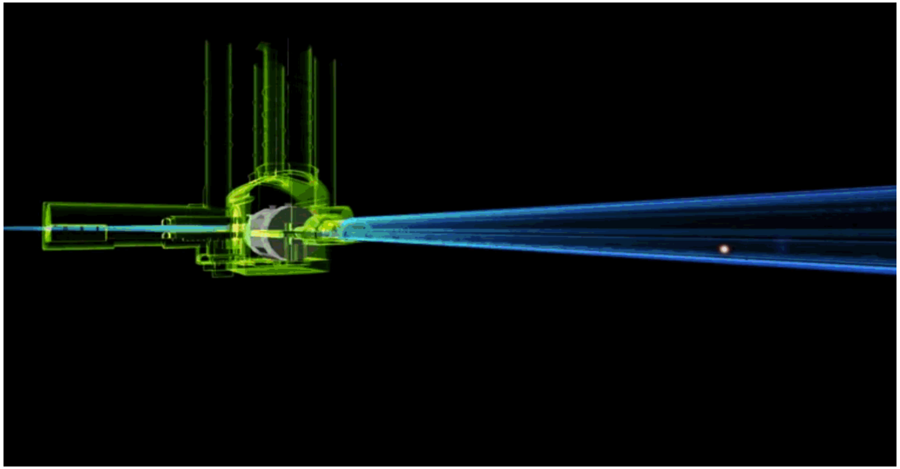
Computational topography

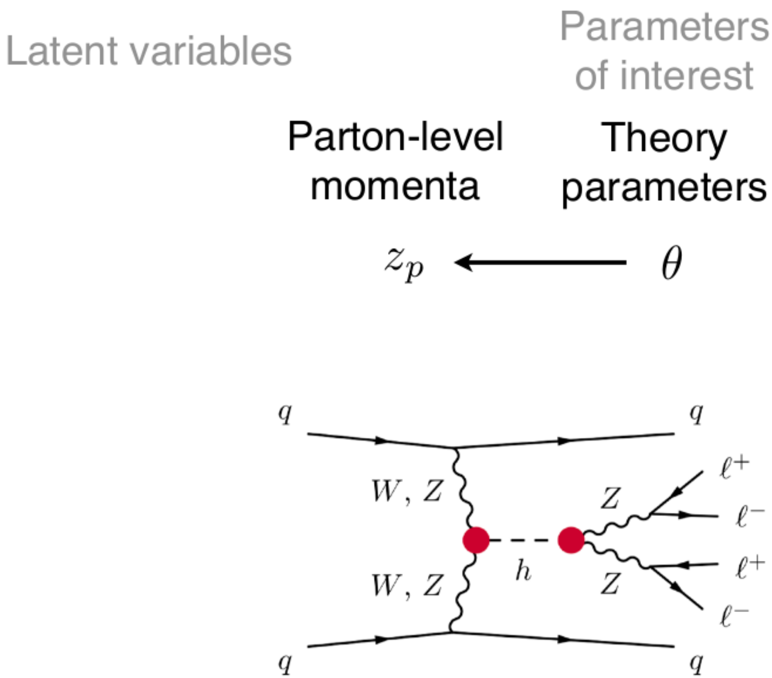
Astronomy

Particle physics

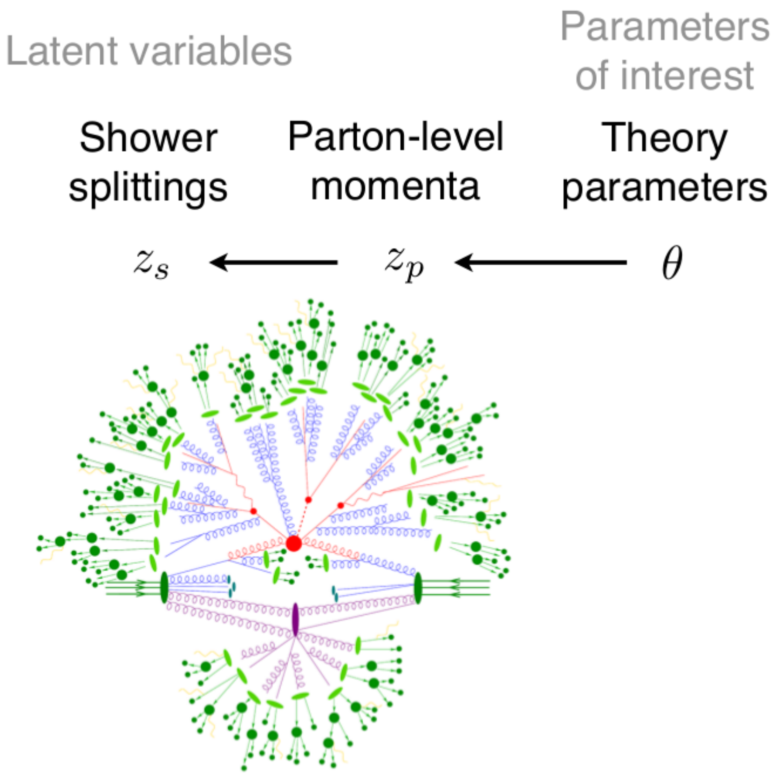


$$
\begin{gathered}
\mathcal{L}_{SM} = -\tfrac{1}{2}\partial_\nu g^a_\mu \partial_\nu g^a_\mu - g_s f^{abc}\partial_\mu g^a_\nu g^b_\mu g^c_\nu - \tfrac{1}{4}g_s^2 f^{abc} f^{ade} g^b_\mu g^c_\nu g^d_\mu g^e_\nu - \partial_\nu W^+_\mu \partial_\nu W^-_\mu - \\
M^2 W^+_\mu W^-_\mu - \tfrac{1}{2}\partial_\nu Z^0_\mu \partial_\nu Z^0_\mu - \tfrac{1}{2c_w^2} M^2 Z^0_\mu Z^0_\mu - \tfrac{1}{2}\partial_\mu A_\nu \partial_\mu A_\nu - igc_w(\partial_\nu Z^0_\mu (W^+_\mu W^-_\nu - \\
W^+_\nu W^-_\mu) - Z^0_\nu (W^+_\mu \partial_\nu W^-_\mu - W^-_\mu \partial_\nu W^+_\mu) + Z^0_\mu (W^+_\nu \partial_\nu W^-_\mu - W^-_\nu \partial_\nu W^+_\mu)) - \\
igs_w(\partial_\nu A_\mu (W^+_\mu W^-_\nu - W^+_\nu W^-_\mu) - A_\nu (W^+_\mu \partial_\nu W^-_\mu - W^-_\mu \partial_\nu W^+_\mu) + A_\mu (W^+_\nu \partial_\nu W^-_\mu - \\
W^-_\nu \partial_\nu W^+_\mu)) - \tfrac{1}{2}g^2 W^+_\mu W^-_\mu W^+_\nu W^-_\nu + \tfrac{1}{2}g^2 W^+_\mu W^-_\nu W^+_\mu W^-_\nu + g^2 c_w^2 (Z^0_\mu W^+_\mu Z^0_\nu W^-_\nu - \\
Z^0_\mu Z^0_\mu W^+_\nu W^-_\nu) + g^2 s_w^2 (A_\mu W^+_\mu A_\nu W^-_\nu - A_\mu A_\mu W^+_\nu W^-_\nu) + g^2 s_w c_w (A_\mu Z^0_\nu (W^+_\mu W^-_\nu - \\
W^+_\nu W^-_\mu) - 2A_\mu Z^0_\mu W^+_\nu W^-_\nu) - \tfrac{1}{2}\partial_\mu H \partial_\mu H - 2M^2 \alpha_h H^2 - \partial_\mu \phi^+ \partial_\mu \phi^- - \tfrac{1}{2}\partial_\mu \phi^0 \partial_\mu \phi^0 - \\
\beta_h \left(\tfrac{2M^2}{g^2} + \tfrac{2M}{g} H + \tfrac{1}{2}(H^2 + \phi^0\phi^0 + 2\phi^+\phi^-)\right) + \tfrac{2M^4}{g^2}\alpha_h - \\
g\alpha_h M \left(H^3 + H\phi^0\phi^0 + 2H\phi^+\phi^-\right) - \\
\tfrac{1}{8}g^2 \alpha_h \left(H^4 + (\phi^0)^4 + 4(\phi^+\phi^-)^2 + 4(\phi^0)^2\phi^+\phi^- + 4H^2\phi^+\phi^- + 2(\phi^0)^2 H^2\right) - \\
gMW^+_\mu W^-_\mu H - \tfrac{1}{2}g\tfrac{M}{c_w^2} Z^0_\mu Z^0_\mu H - \\
\tfrac{1}{2}ig\left(W^+_\mu(\phi^0\partial_\mu\phi^- - \phi^-\partial_\mu\phi^0) - W^-_\mu(\phi^0\partial_\mu\phi^+ - \phi^+\partial_\mu\phi^0)\right) + \\
\tfrac{1}{2}g\left(W^+_\mu(H\partial_\mu\phi^- - \phi^-\partial_\mu H) + W^-_\mu(H\partial_\mu\phi^+ - \phi^+\partial_\mu H)\right) + \tfrac{1}{2}g\tfrac{1}{c_w}(Z^0_\mu(H\partial_\mu\phi^0 - \phi^0\partial_\mu H) + \\
M\left(\tfrac{1}{c_w} Z^0_\mu \partial_\mu \phi^0 + W^+_\mu \partial_\mu \phi^- + W^-_\mu \partial_\mu \phi^+\right) - ig\tfrac{s_w^2}{c_w} M Z^0_\mu (W^+_\mu \phi^- - W^-_\mu \phi^+) + igs_w M A_\mu (W^+_\mu \phi^- - \\
W^-_\mu \phi^+) - ig\tfrac{1-2c_w^2}{2c_w} Z^0_\mu (\phi^+\partial_\mu\phi^- - \phi^-\partial_\mu\phi^+) + igs_w A_\mu (\phi^+\partial_\mu\phi^- - \phi^-\partial_\mu\phi^+) - \\
\tfrac{1}{4}g^2 W^+_\mu W^-_\mu \left(H^2 + (\phi^0)^2 + 2\phi^+\phi^-\right) - \tfrac{1}{8}g^2 \tfrac{1}{c_w^2} Z^0_\mu Z^0_\mu \left(H^2 + (\phi^0)^2 + 2(2s_w^2 - 1)^2 \phi^+\phi^-\right) - \\
\tfrac{1}{2}g^2 \tfrac{s_w^2}{c_w} Z^0_\mu \phi^0 (W^+_\mu \phi^- + W^-_\mu \phi^+) - \tfrac{1}{2}ig^2 \tfrac{s_w^2}{c_w} Z^0_\mu H (W^+_\mu \phi^- - W^-_\mu \phi^+) + \tfrac{1}{2}g^2 s_w A_\mu \phi^0 (W^+_\mu \phi^- + \\
W^-_\mu \phi^+) + \tfrac{1}{2}ig^2 s_w A_\mu H (W^+_\mu \phi^- - W^-_\mu \phi^+) - g^2 \tfrac{s_w}{c_w}(2c_w^2 - 1) Z^0_\mu A_\mu \phi^+\phi^- - \\
g^2 s_w^2 A_\mu A_\mu \phi^+\phi^- + \tfrac{1}{2}ig_s \lambda^a_{ij}(\bar{q}^\sigma_i \gamma^\mu q^\sigma_j) g^a_\mu - \bar{e}^\lambda(\gamma\partial + m^\lambda_e)e^\lambda - \bar{\nu}^\lambda(\gamma\partial + m^\lambda_\nu)\nu^\lambda - \bar{u}^\lambda_j(\gamma\partial + \\
m^\lambda_u)u^\lambda_j - \bar{d}^\lambda_j(\gamma\partial + m^\lambda_d)d^\lambda_j + igs_w A_\mu \left(-(\bar{e}^\lambda\gamma^\mu e^\lambda) + \tfrac{2}{3}(\bar{u}^\lambda_j\gamma^\mu u^\lambda_j) - \tfrac{1}{3}(\bar{d}^\lambda_j\gamma^\mu d^\lambda_j)\right) + \\
\tfrac{ig}{4c_w} Z^0_\mu \{(\bar{\nu}^\lambda\gamma^\mu(1+\gamma^5)\nu^\lambda) + (\bar{e}^\lambda\gamma^\mu(4s_w^2 - 1 - \gamma^5)e^\lambda) + (\bar{d}^\lambda_j\gamma^\mu(\tfrac{4}{3}s_w^2 - 1 - \gamma^5)d^\lambda_j) + \\
(\bar{u}^\lambda_j\gamma^\mu(1 - \tfrac{8}{3}s_w^2 + \gamma^5)u^\lambda_j)\} + \tfrac{ig}{2\sqrt{2}} W^+_\mu \left((\bar{\nu}^\lambda\gamma^\mu(1+\gamma^5)U^{lep}{}_{\lambda\kappa} e^\kappa) + (\bar{u}^\lambda_j\gamma^\mu(1+\gamma^5)C_{\lambda\kappa} d^\kappa_j)\right) + \\
\tfrac{ig}{2\sqrt{2}} W^-_\mu \left((\bar{e}^\kappa U^{lep\dagger}{}_{\kappa\lambda}\gamma^\mu(1+\gamma^5)\nu^\lambda) + (\bar{d}^\kappa_j C^\dagger_{\kappa\lambda}\gamma^\mu(1+\gamma^5)u^\lambda_j)\right) + \\
\tfrac{ig}{2M\sqrt{2}}\phi^+ \left(-m^\kappa_e(\bar{\nu}^\lambda U^{lep}{}_{\lambda\kappa}(1-\gamma^5)e^\kappa) + m^\lambda_\nu(\bar{\nu}^\lambda U^{lep}{}_{\lambda\kappa}(1+\gamma^5)e^\kappa\right) + \\
\tfrac{ig}{2M\sqrt{2}}\phi^- \left(m^\lambda_e(\bar{e}^\lambda U^{lep\dagger}{}_{\lambda\kappa}(1+\gamma^5)\nu^\kappa) - m^\kappa_\nu(\bar{e}^\lambda U^{lep\dagger}{}_{\lambda\kappa}(1-\gamma^5)\nu^\kappa\right) - \tfrac{g}{2}\tfrac{m^\lambda_\nu}{M} H(\bar{\nu}^\lambda\nu^\lambda) - \\
\tfrac{g}{2}\tfrac{m^\lambda_e}{M} H(\bar{e}^\lambda e^\lambda) + \tfrac{ig}{2}\tfrac{m^\lambda_\nu}{M}\phi^0(\bar{\nu}^\lambda\gamma^5\nu^\lambda) - \tfrac{ig}{2}\tfrac{m^\lambda_e}{M}\phi^0(\bar{e}^\lambda\gamma^5 e^\lambda) - \tfrac{1}{4}\bar{\nu}_\lambda M^R_{\lambda\kappa}(1-\gamma_5)\hat{\nu}_\kappa - \\
\tfrac{1}{4}\overline{\bar{\nu}_\lambda M^R_{\lambda\kappa}(1-\gamma_5)\hat{\nu}_\kappa} + \tfrac{ig}{2M\sqrt{2}}\phi^+ \left(-m^\kappa_d(\bar{u}^\lambda_j C_{\lambda\kappa}(1-\gamma^5)d^\kappa_j) + m^\lambda_u(\bar{u}^\lambda_j C_{\lambda\kappa}(1+\gamma^5)d^\kappa_j\right) + \\
\tfrac{ig}{2M\sqrt{2}}\phi^- \left(m^\lambda_d(\bar{d}^\lambda_j C^\dagger_{\lambda\kappa}(1+\gamma^5)u^\kappa_j) - m^\kappa_u(\bar{d}^\lambda_j C^\dagger_{\lambda\kappa}(1-\gamma^5)u^\kappa_j\right) - \tfrac{g}{2}\tfrac{m^\lambda_u}{M} H(\bar{u}^\lambda_j u^\lambda_j) - \\
\tfrac{g}{2}\tfrac{m^\lambda_d}{M} H(\bar{d}^\lambda_j d^\lambda_j) + \tfrac{ig}{2}\tfrac{m^\lambda_u}{M}\phi^0(\bar{u}^\lambda_j\gamma^5 u^\lambda_j) - \tfrac{ig}{2}\tfrac{m^\lambda_d}{M}\phi^0(\bar{d}^\lambda_j\gamma^5 d^\lambda_j) + \bar{G}^a\partial^2 G^a + g_s f^{abc}\partial_\mu \bar{G}^a G^b g^c_\mu + \\
\bar{X}^+(\partial^2 - M^2)X^+ + \bar{X}^-(\partial^2 - M^2)X^- + \bar{X}^0(\partial^2 - \tfrac{M^2}{c_w^2})X^0 + \bar{Y}\partial^2 Y + igc_w W^+_\mu(\partial_\mu \bar{X}^0 X^- - \\
\partial_\mu \bar{X}^+ X^0) + igs_w W^+_\mu(\partial_\mu \bar{Y} X^- - \partial_\mu \bar{X}^+ Y) + igc_w W^-_\mu(\partial_\mu \bar{X}^- X^0 - \\
\partial_\mu \bar{X}^0 X^+) + igs_w W^-_\mu(\partial_\mu \bar{X}^- Y - \partial_\mu \bar{Y} X^+) + igc_w Z^0_\mu(\partial_\mu \bar{X}^+ X^+ - \\
\partial_\mu \bar{X}^- X^-) + igs_w A_\mu(\partial_\mu \bar{X}^+ X^+ - \\
\partial_\mu \bar{X}^- X^-) - \tfrac{1}{2}gM\left(\bar{X}^+ X^+ H + \bar{X}^- X^- H + \tfrac{1}{c_w^2}\bar{X}^0 X^0 H\right) + \tfrac{1-2c_w^2}{2c_w} igM\left(\bar{X}^+ X^0 \phi^+ - \bar{X}^- X^0 \phi^-\right) + \\
\tfrac{1}{2c_w} igM\left(\bar{X}^0 X^- \phi^+ - \bar{X}^0 X^+ \phi^-\right) + igMs_w\left(\bar{X}^0 X^- \phi^+ - \bar{X}^0 X^+ \phi^-\right) + \\
\tfrac{1}{2} igM\left(\bar{X}^+ X^+ \phi^0 - \bar{X}^- X^- \phi^0\right) .
\end{gathered}
$$





Credits: Johann Brehmer.


Latent variables
Parameters of interest
Shower splittings
Parton-level momenta
Theory parameters
$z_s \leftarrow z_p \leftarrow \theta$

Latent variables

Parameters of interest

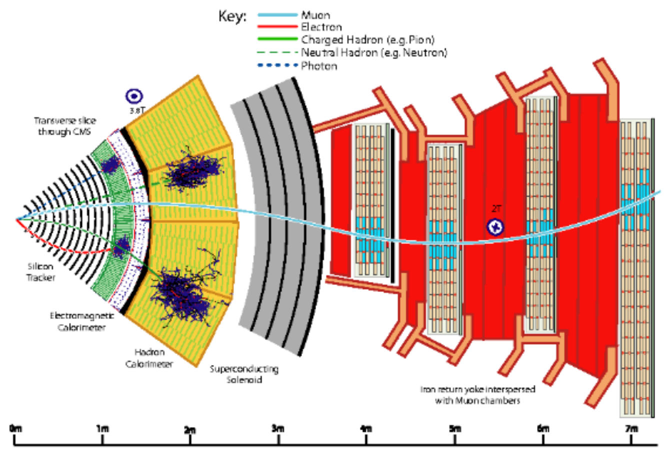
Detector interactions

Shower splittings

Parton-level momenta

Theory parameters

$$z_d \longleftarrow z_s \longleftarrow z_p \longleftarrow \theta$$

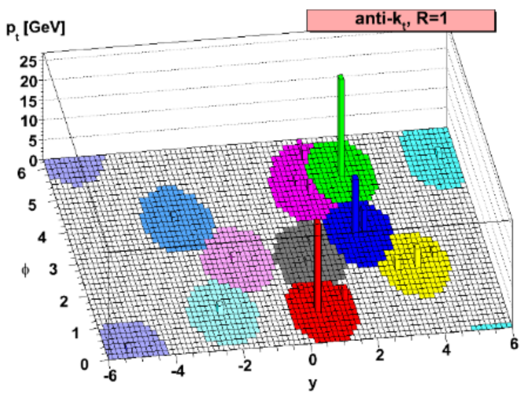


Credits: Johann Brehmer.

Features | Latent variables | Parameters of interest

Observables | Detector interactions | Shower splittings | Parton-level momenta | Theory parameters

$$x \leftarrow z_d \leftarrow z_s \leftarrow z_p \leftarrow \theta$$



[Image source: M. Cacciari, G. Salam, G. Soyez 0802.1189]

Credits: Johann Brehmer.

$$p(x|\theta) = \underbrace{\iiint}_{\text{intractable}} p(z_p|\theta)p(z_s|z_p)p(z_d|z_s)p(x|z_d)dz_p dz_s dz_d$$

Likelihood ratio

The likelihood ratio

$$r(x|\theta_0, \theta_1) = \frac{p(x|\theta_0)}{p(x|\theta_1)}$$

is the quantity that is central to many statistical inference procedures.

Examples

- Frequentist hypothesis testing
- Supervised learning
- Bayesian posterior sampling with MCMC
- Bayesian posterior inference through Variational Inference
- Generative adversarial networks
- Empirical Bayes with Adversarial Variational Optimization

The likelihood $p(x|\theta)$ is actually rarely needed.

When solving a problem of interest, do not solve a more general problem as an intermediate step. – Vladimir Vapnik

$$\frac{p_{\mathbf{X}}(\mathbf{x}|\theta_0)}{p_{\mathbf{X}}(\mathbf{x}|\theta_1)} = r(\mathbf{x}; \theta_0, \theta_1)$$

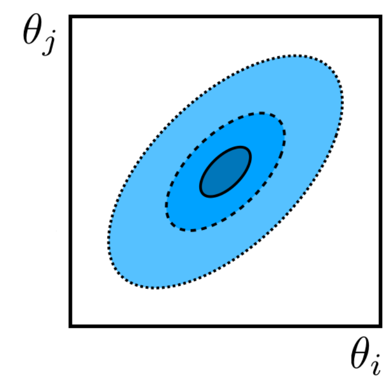
Direct likelihood ratio estimation is simpler than density estimation.
(This is fortunate, we are in the likelihood-free scenario!)

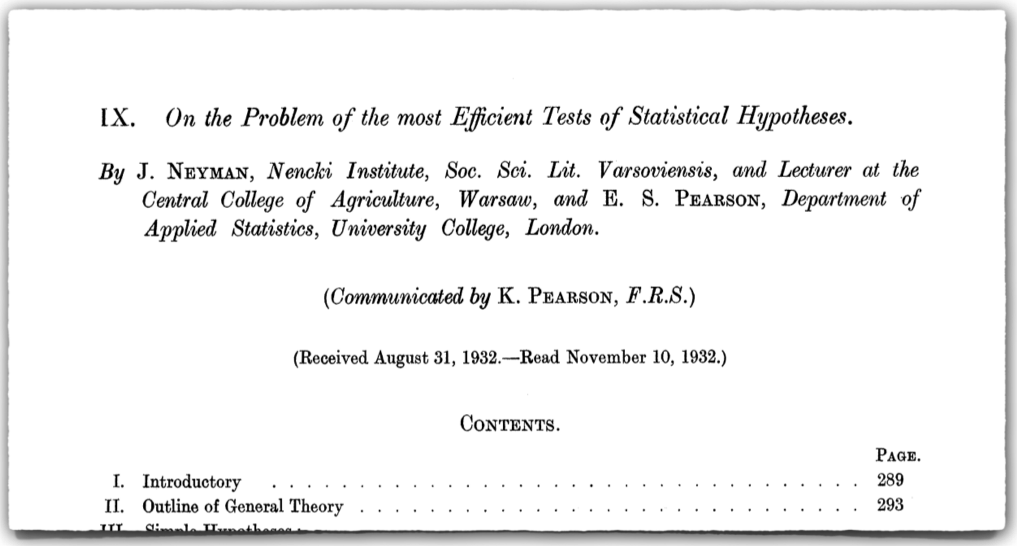
The frequentist physicist's way

The Neyman-Pearson lemma states that the likelihood ratio

$$r(x|\theta_0, \theta_1) = \frac{p(x|\theta_0)}{p(x|\theta_1)}$$

is the most powerful test statistic to discriminate between a null hypothesis θ_0 and an alternative θ_1.





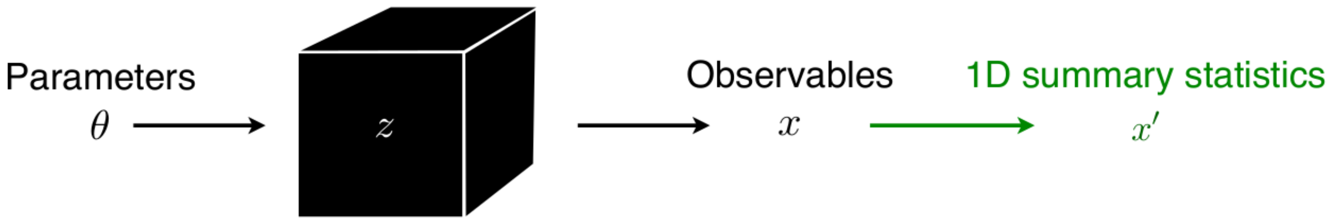
IX. *On the Problem of the most Efficient Tests of Statistical Hypotheses.*

By J. Neyman, *Nencki Institute, Soc. Sci. Lit. Varsoviensis, and Lecturer at the Central College of Agriculture, Warsaw, and* E. S. Pearson, *Department of Applied Statistics, University College, London.*

(*Communicated by* K. Pearson, *F.R.S.*)

(Received August 31, 1932.—Read November 10, 1932.)

Contents.



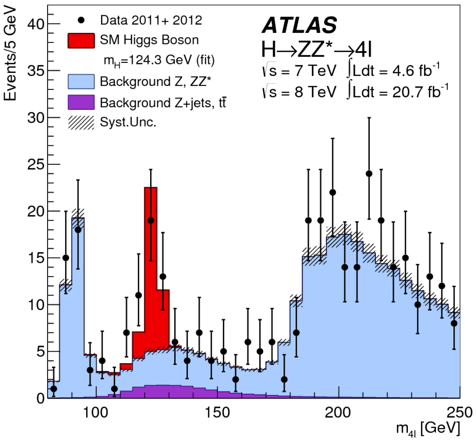
Define a projection function $s : \mathcal{X} \rightarrow \mathbb{R}$ mapping observables x to a summary statistics $x' = s(x)$.

Then, approximate the likelihood $p(x|\theta)$ as

$$p(x|\theta) \approx \hat{p}(x|\theta) = p(x'|\theta).$$

From this it comes

$$\frac{p(x|\theta_0)}{p(x|\theta_1)} \approx \frac{\hat{p}(x|\theta_0)}{\hat{p}(x|\theta_1)} = \hat{r}(x|\theta_0, \theta_1).$$

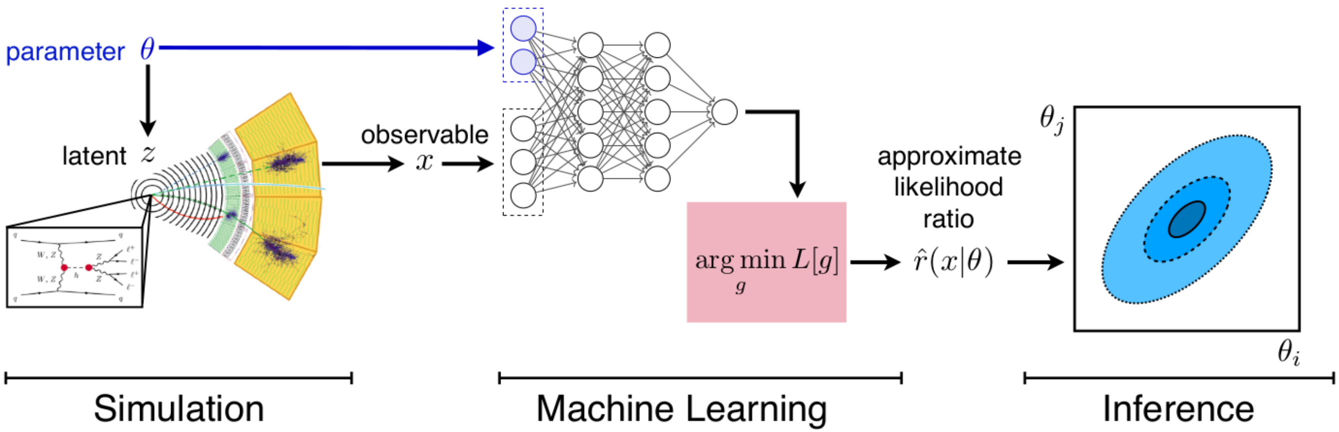


CARL

Supervised learning provides a way to automatically construct s:

- Let us consider a binary classifier $\hat{s}$ (e.g., a neural network) trained to distinguish $x \sim p(x|\theta_0)$ from $x \sim p(x|\theta_1)$.
- $\hat{s}$ is trained by minimizing the cross-entropy loss

$$\begin{aligned} L_{XE}[\hat{s}] = -\mathbb{E}_{p(x|\theta)\pi(\theta)}[&1(\theta = \theta_0) \log \hat{s}(x) + \\ &1(\theta = \theta_1) \log(1 - \hat{s}(x))] \end{aligned}$$



Reference: Cranmer et al, 2015 [arXiv:1506.02169].

The solution $\hat{s}$ found after training approximates the optimal classifier

$$\hat{s}(x) \approx s^*(x) = \frac{p(x|\theta_1)}{p(x|\theta_0) + p(x|\theta_1)}.$$

Therefore,

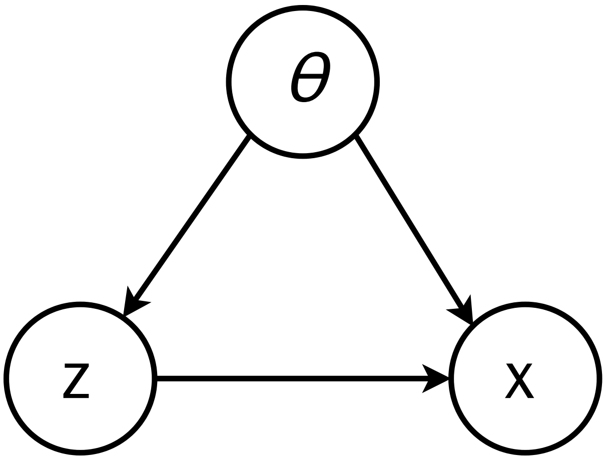
$$r(x|\theta_0, \theta_1) \approx \hat{r}(x|\theta_0, \theta_1) = \frac{1 - \hat{s}(x)}{\hat{s}(x)}$$

That is, supervised classification is equivalent to likelihood ratio estimation.

Bayesian inference

For a given model $p(x, z, \theta)$,
Bayesian inference usually consists in computing the posterior

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}.$$



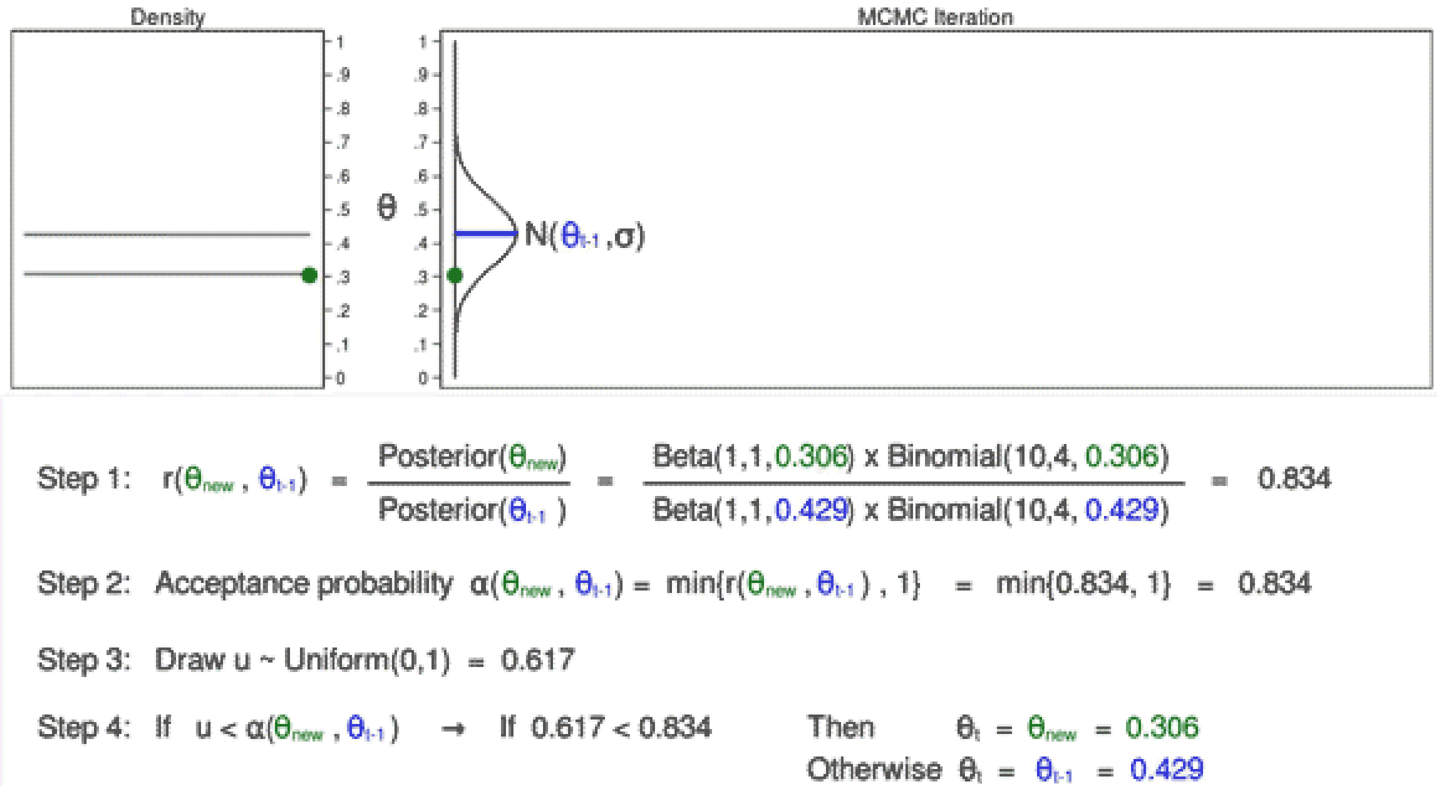
For most cases, this is intractable since it requires evaluating the evidence

$$p(x) = \int p(x|\theta)p(\theta)d\theta.$$

In the likelihood-free scenario, this is even less tractable since we cannot even evaluate the likelihood

$$p(x|\theta) = \int p(x, z|\theta)dz.$$

Posterior sampling



Credits: Chuck Huber, 2016.

Likelihood-free MCMC with Approximate Likelihood Ratios

Joeri Hermans
joeri.hermans@doct.uliege.be
University of Liège, Belgium

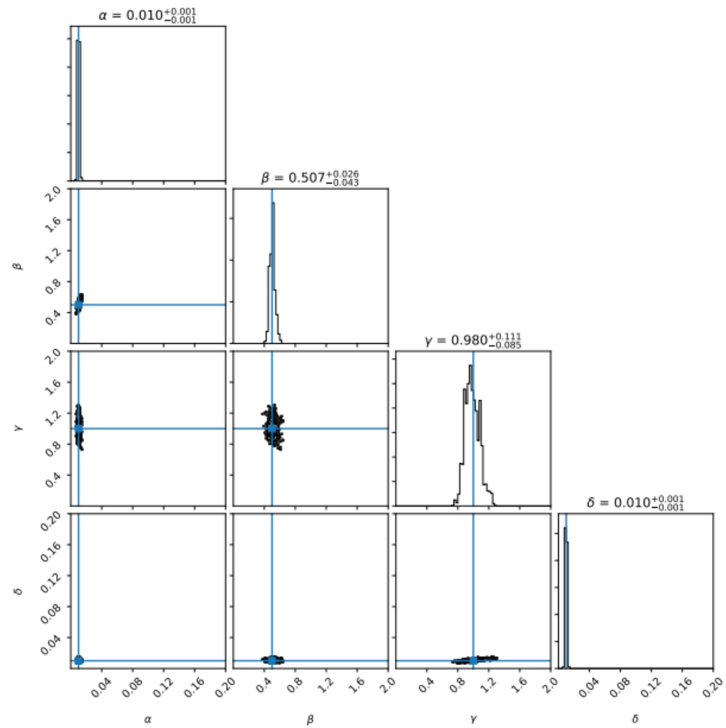
Volodimir Begy
volodimir.begy@cern.ch
University of Vienna, Austria
CERN, Switzerland

Gilles Louppe
g.louppe@uliege.be
University of Liège, Belgium

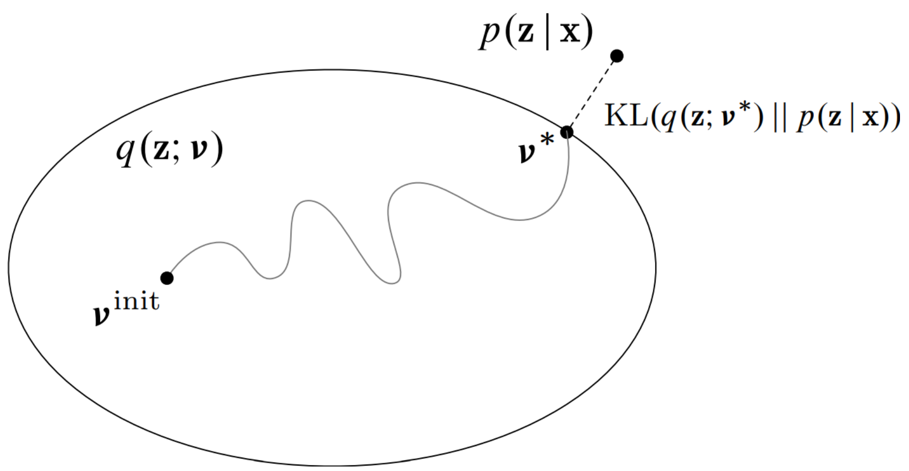
Algorithm 2 Likelihood-free Metropolis-Hastings

Inputs:	Initial parameter $\boldsymbol{\theta}_0$.
	Transition distribution $q(\boldsymbol{\theta})$.
	Parameterized classifier $\mathbf{s}(\mathbf{x}, \boldsymbol{\theta})$.
	Observations $\mathcal{O}$.
Outputs:	Markov chain $\boldsymbol{\theta}_{0:n}$
Hyperparameters:	Steps n.

$t \leftarrow 0$
$\boldsymbol{\theta}_t \leftarrow \boldsymbol{\theta}_0$
for $t < n$ **do**
  $\boldsymbol{\theta}' \sim q(\boldsymbol{\theta} \mid \boldsymbol{\theta}_t)$
  $\lambda \leftarrow \sum_{\mathbf{x}\in\mathcal{O}} \log \hat{r}_e(\mathbf{x}, \boldsymbol{\theta}') - \sum_{\mathbf{x}\in\mathcal{O}} \log \hat{r}_e(\mathbf{x}, \boldsymbol{\theta}_t)$
  $\rho \leftarrow \min\left\{\exp(\lambda)\dfrac{q(\boldsymbol{\theta}_t|\boldsymbol{\theta}')}{q(\boldsymbol{\theta}'|\boldsymbol{\theta}_t)},\, 1\right\}$
  $\boldsymbol{\theta}_{t+1} \leftarrow \begin{cases}\boldsymbol{\theta}' & \text{with probability } \rho \\ \boldsymbol{\theta}_t & \text{with probability } 1-\rho\end{cases}$
  $t \leftarrow t + 1$
end for
return $\boldsymbol{\theta}_{0:n}$



Reference: Hermans et al, 2019 [arXiv:1903.04057].



Likelihood-free Variational inference

Let $q(\mathbf{x}_n)$ be the empirical distribution on the observations $\mathbf{x}$ and consider using it in a "variational joint" $q(\mathbf{x}_n, \mathbf{z}_n \mid \boldsymbol{\beta}) = q(\mathbf{x}_n)q(\mathbf{z}_n \mid \mathbf{x}_n, \boldsymbol{\beta})$. Now subtract the log empirical $\log q(\mathbf{x}_n)$ from the ELBO above. The ELBO reduces to

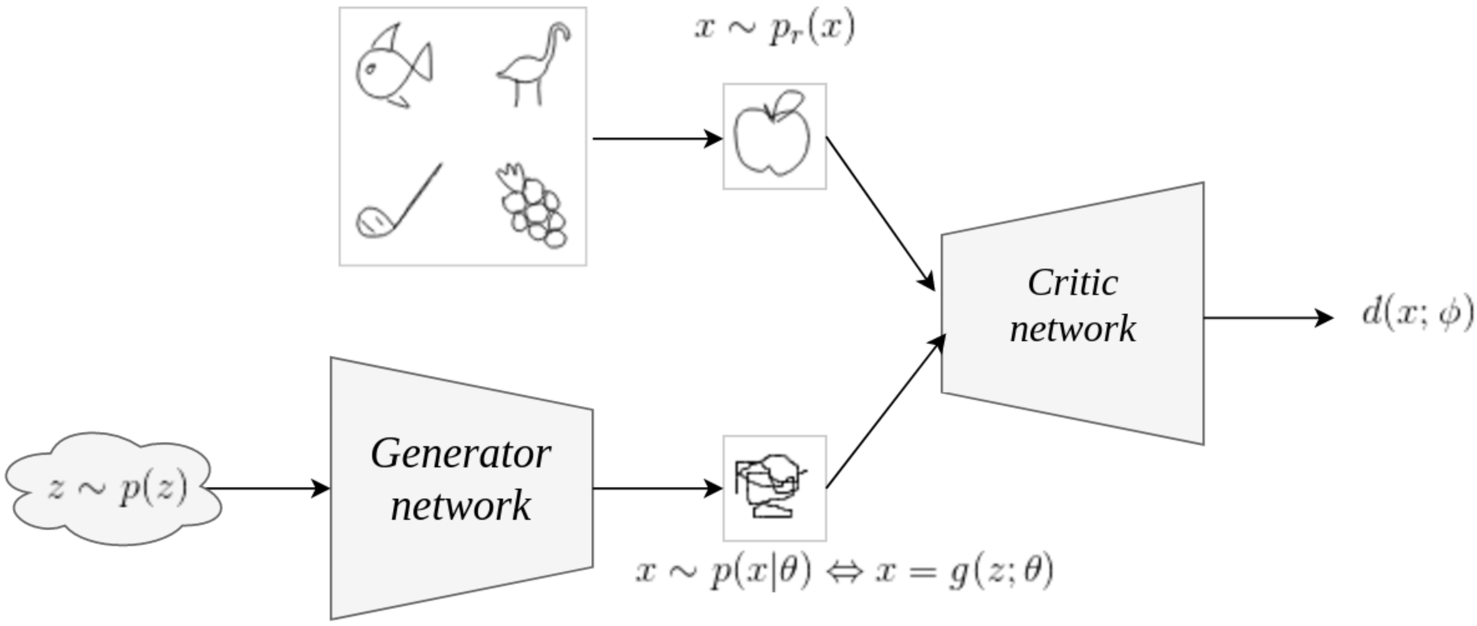
$$\mathcal{L} \propto \mathbb{E}_{q(\boldsymbol{\beta})}[\log p(\boldsymbol{\beta}) - \log q(\boldsymbol{\beta})] + \sum_{n=1}^{N} \mathbb{E}_{q(\boldsymbol{\beta})q(\mathbf{z}_n \mid \mathbf{x}_n, \boldsymbol{\beta})} \left[\log \frac{p(\mathbf{x}_n, \mathbf{z}_n \mid \boldsymbol{\beta})}{q(\mathbf{x}_n, \mathbf{z}_n \mid \boldsymbol{\beta})} \right]. \tag{4}$$

(Here the proportionality symbol means equality up to additive constants.) Thus the ELBO is a function of the ratio of two intractable densities. If we can form an estimator of this ratio, we can proceed with optimizing the ELBO.

We apply techniques for ratio estimation [51]. It is a key idea in GANs [37, 56], and similar ideas have rearisen in statistics and physics [21, 8].

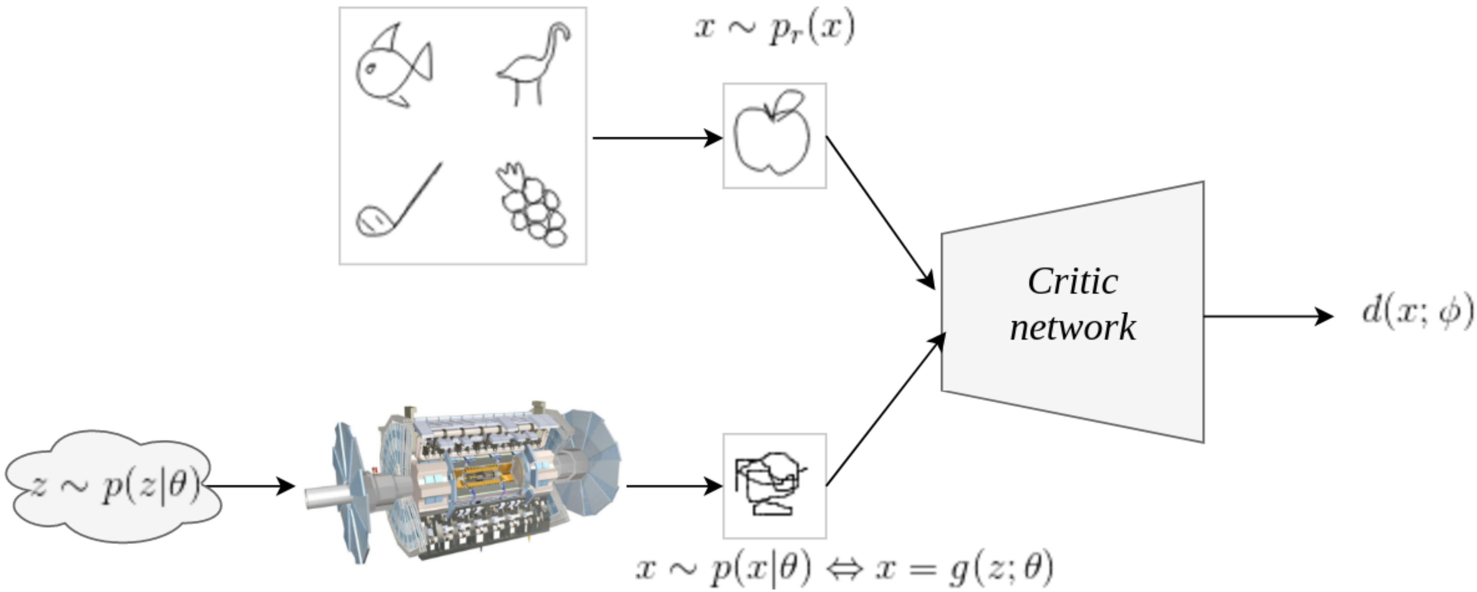
Reference: Tran et al, 2017 [arXiv:1702.08896].

Generative adversarial networks



Reference: Goodfellow et al, 2014 [arXiv:1406.2661].

Adversarial Variational Optimization



Replace g with an actual scientific simulator!

Reference: Louppe et al, 2017 [arXiv:1707.07113].

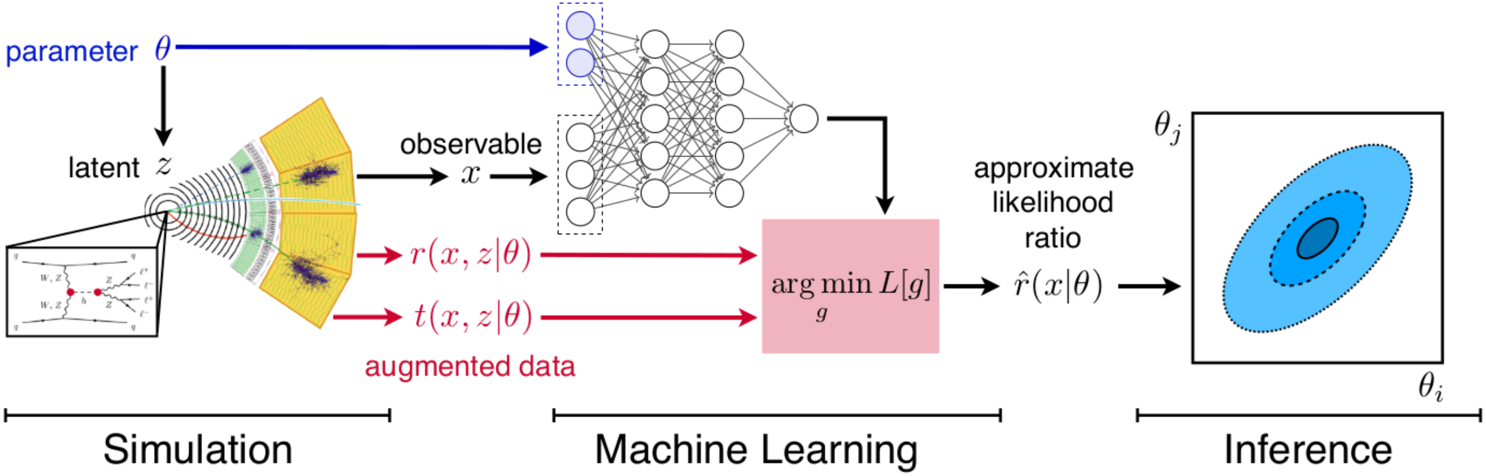
Key insights

- Replace the generative network with a non-differentiable forward simulator $g(z; \theta)$.
- Let the neural network critic figure out how to adjust the simulator parameters.
- Combine with variational optimization to bypass the non-differentiability by optimizing upper bounds of the adversarial objectives

$$U_d(\phi) = \mathbb{E}_{\theta \sim q(\theta;\psi)} \left[\mathcal{L}_d(\phi)\right]$$
$$U_g(\psi) = \mathbb{E}_{\theta \sim q(\theta;\psi)} \left[\mathcal{L}_g(\theta)\right]$$

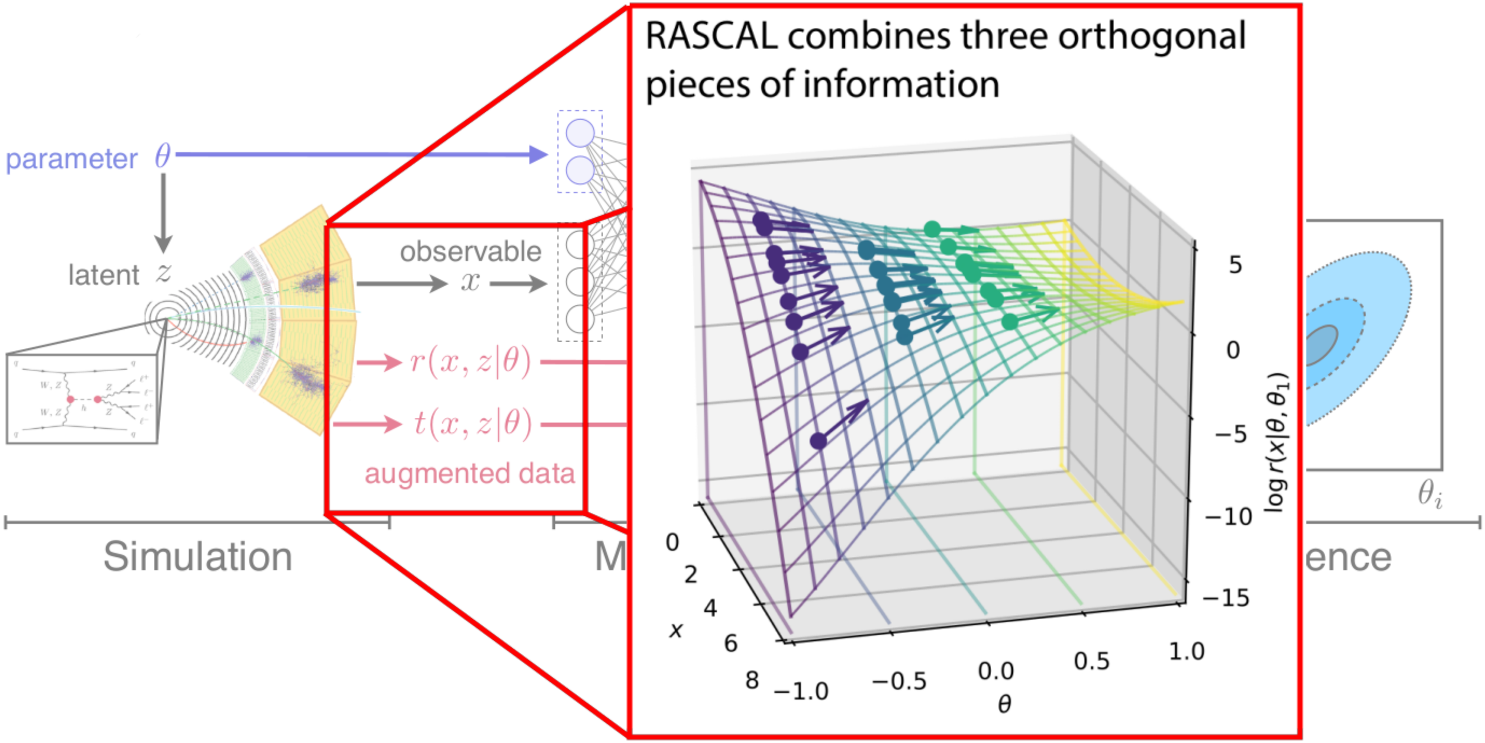
  respectively over ϕ and ψ.
- Effectively, this amounts to empirical Bayes guided by the likelihood ratios estimated from the critic.

Mining gold

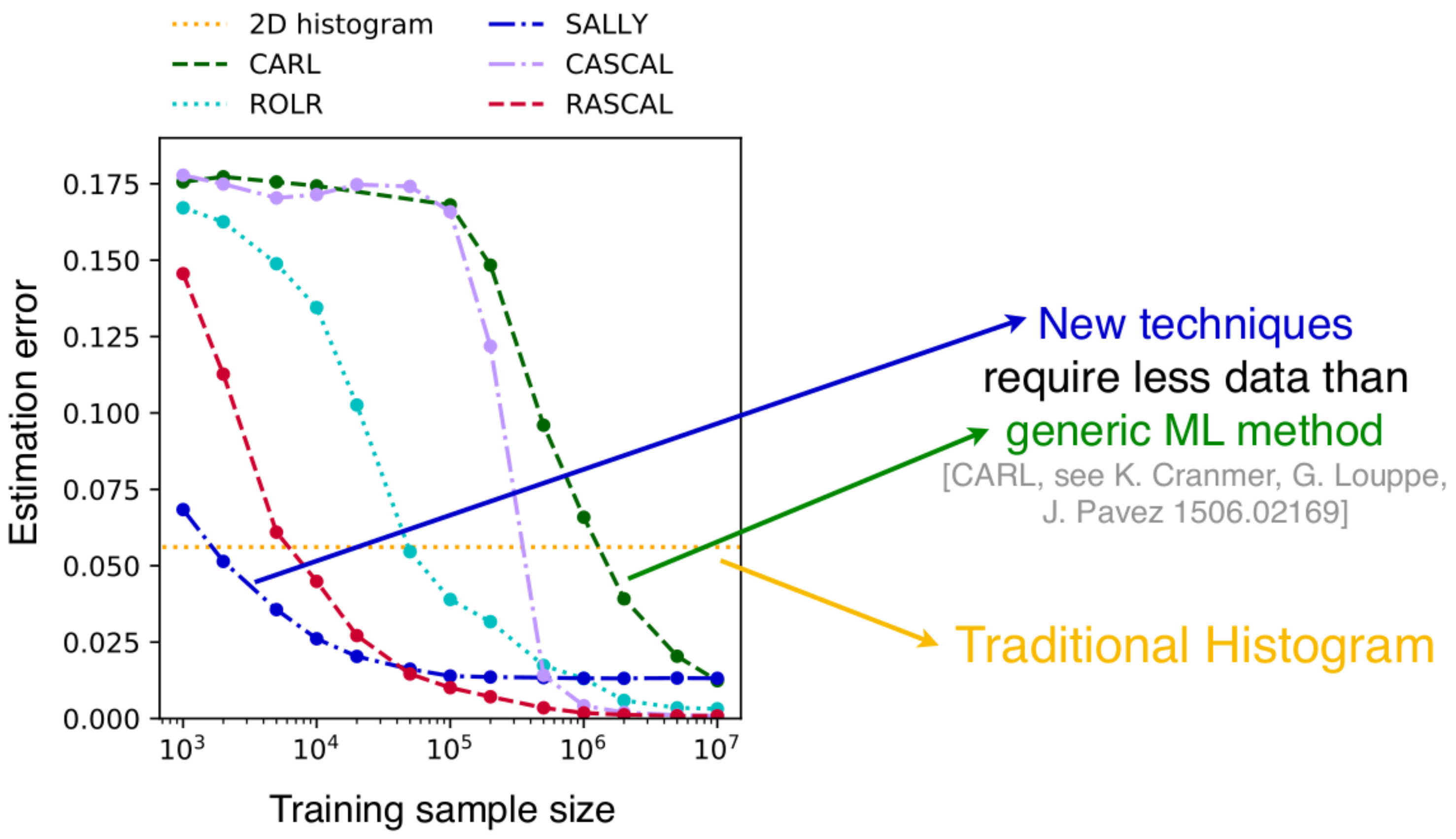


Reference: Brehmer et al, 2018 [arXiv:1805.12244].

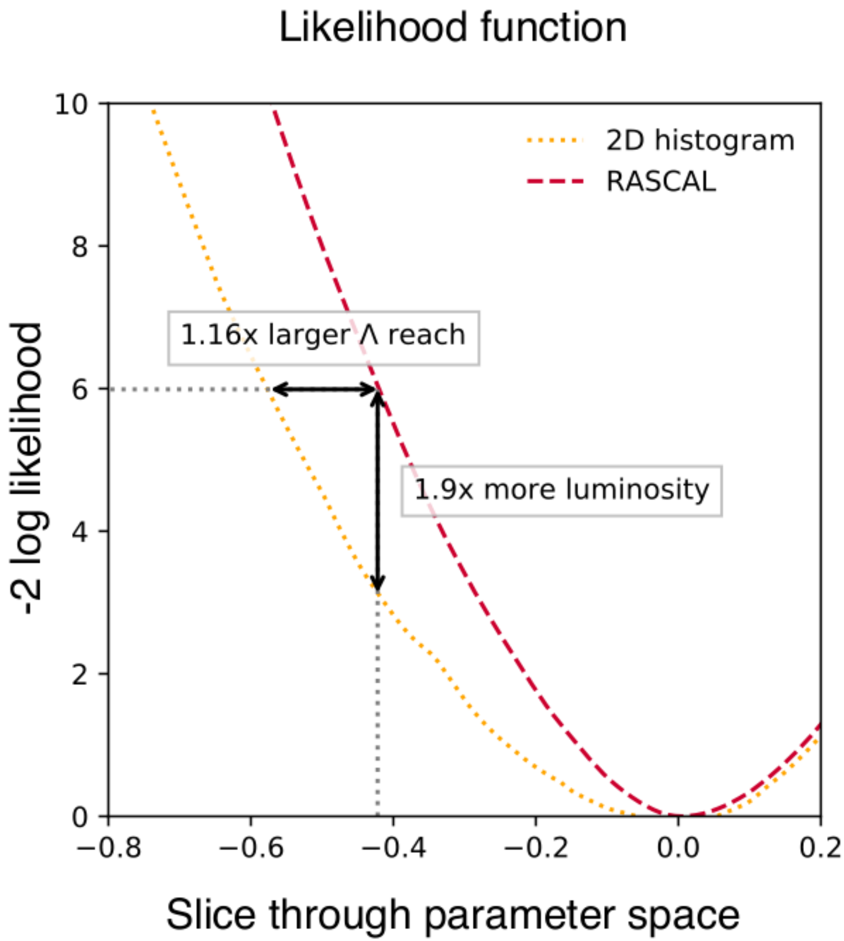
Mining gold


RASCAL combines three orthogonal pieces of information
parameter θ
latent z
observable
x
$r(x,z|\theta)$
$t(x,z|\theta)$
augmented data
Simulation
M
ence
θ_i
$\log r(x|\theta,\theta_1)$
5
0
−5
−10
−15
x
0
2
4
6
8
−1.0
−0.5
0.0
0.5
1.0
θ

Increased data efficiency



Better sensitivity



36 events, assuming SM

Summary

- Much of modern science is based on "likelihood-free" simulations.
- The likelihood-ratio is central to many statistical inference procedures.
- Supervised learning enables likelihood-ratio estimation.
- Better likelihood-ratio estimates can be achieved by mining simulators.

Collaborators




Kyle Cranmer, Juan Pavez, Johann Brehmer, Joeri Hermans

The end.

References

- Hermans, J., Begy, V., & Louppe, G. (2019). Likelihood-free MCMC with Approximate Likelihood Ratios. arXiv preprint arXiv:1903.04057.
- Stoye, M., Brehmer, J., Louppe, G., Pavez, J., & Cranmer, K. (2018). Likelihood-free inference with an improved cross-entropy estimator. arXiv preprint arXiv:1808.00973.
- Baydin, A. G., Heinrich, L., Bhimji, W., Gram-Hansen, B., Louppe, G., Shao, L., ... & Wood, F. (2018). Efficient Probabilistic Inference in the Quest for Physics Beyond the Standard Model. arXiv preprint arXiv:1807.07706.
- Brehmer, J., Louppe, G., Pavez, J., & Cranmer, K. (2018). Mining gold from implicit models to improve likelihood-free inference. arXiv preprint arXiv:1805.12244.
- Brehmer, J., Cranmer, K., Louppe, G., & Pavez, J. (2018). Constraining Effective Field Theories with Machine Learning. arXiv preprint arXiv:1805.00013.
- Brehmer, J., Cranmer, K., Louppe, G., & Pavez, J. (2018). A Guide to Constraining Effective Field Theories with Machine Learning. arXiv preprint arXiv:1805.00020.
- Casado, M. L., Baydin, A. G., Rubio, D. M., Le, T. A., Wood, F., Heinrich, L., ... & Bhimji, W. (2017). Improvements to Inference Compilation for Probabilistic Programming in Large-Scale Scientific Simulators. arXiv preprint arXiv:1712.07901.
- Louppe, G., Hermans, J., & Cranmer, K. (2017). Adversarial Variational Optimization of Non-Differentiable Simulators. arXiv preprint arXiv:1707.07113.
- Cranmer, K., Pavez, J., & Louppe, G. (2015). Approximating likelihood ratios with calibrated discriminative classifiers. arXiv preprint arXiv:1506.02169.