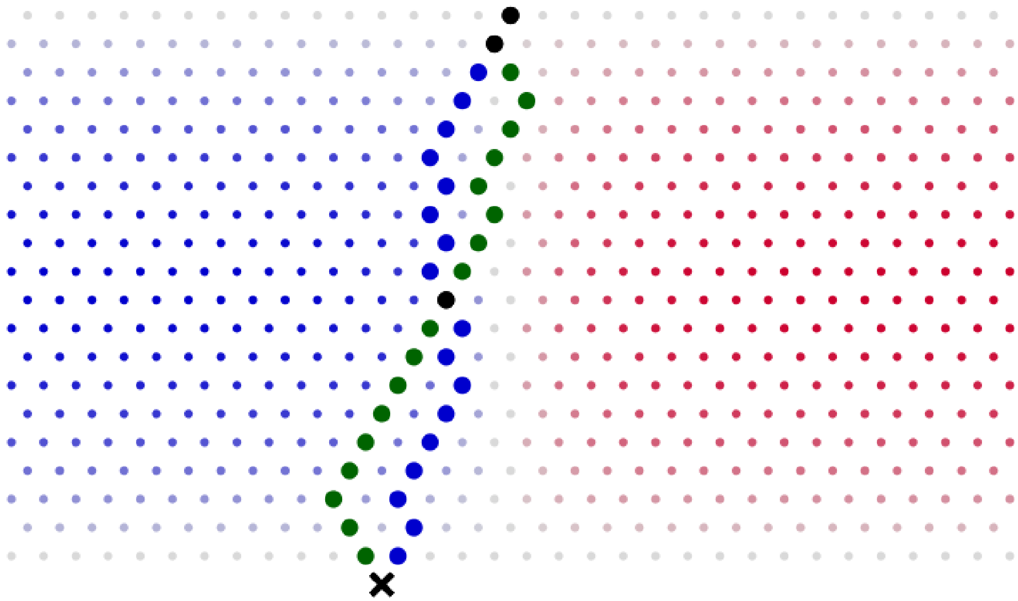
Simulation-based Inference

Likelihood-free MCMC with Amortized Approximate Ratio Estimators (ICML 2020)

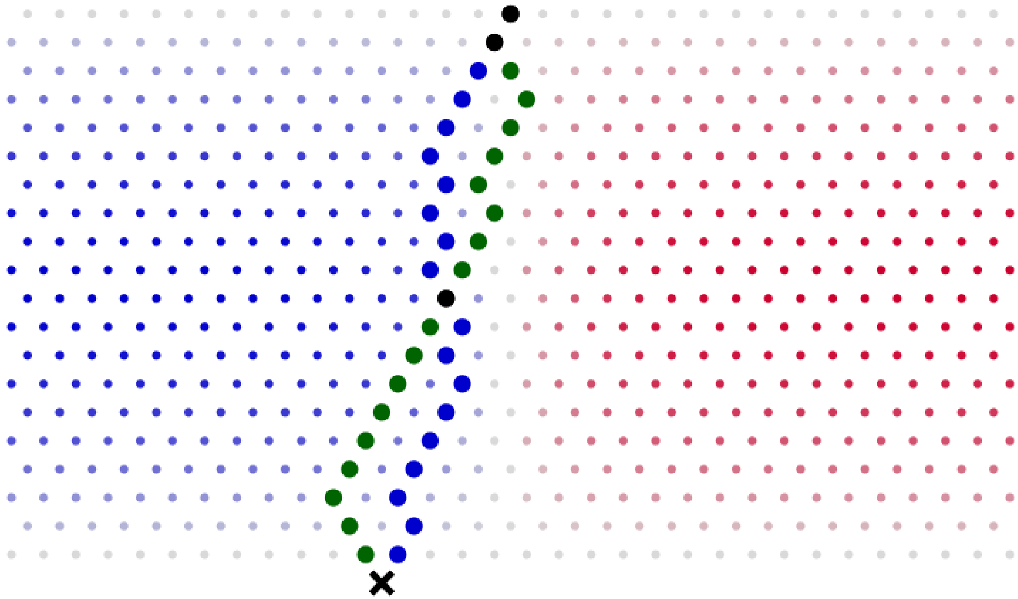
November 17, Parietal

Gilles Louppe
g.louppe@uliege.be

LIÈGE université



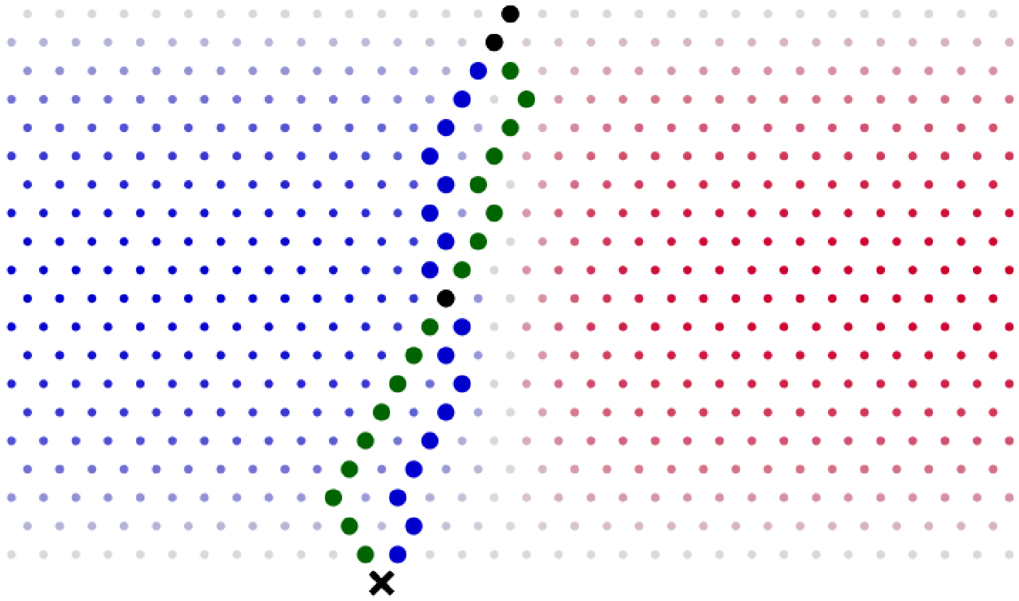
How to estimate the probability θ of going left when hitting a pin?



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

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

Therefore $\hat{\theta} = \arg\max \prod_{x_i} p(x_i|\theta)\pi(\theta)$.

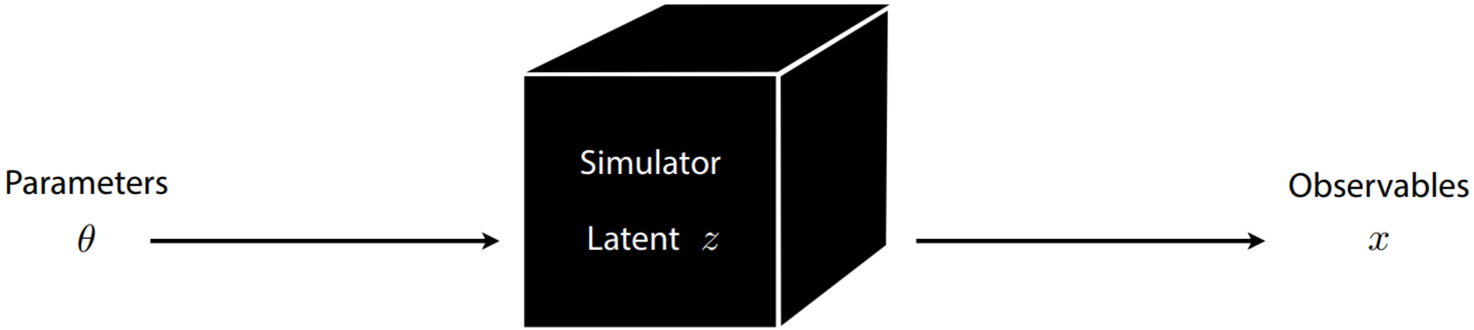


But what if we shift or remove some of the pins?

The Bean machine is a metaphore of simulation-based science:

Bean machine	$\rightarrow$	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.



Prediction:

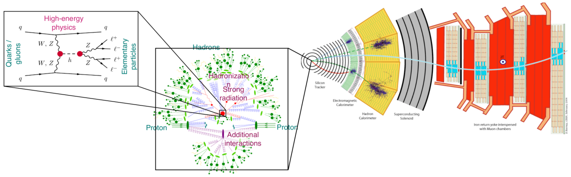
- Well-understood mechanistic model
- Simulator can generate samples

Inference:

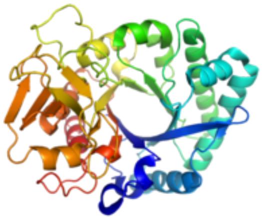
- Likelihood function $p(x|\theta)$ is intractable
- Inference based on estimator $\hat{p}(x|\theta)$

Credits: Johann Brehmer.

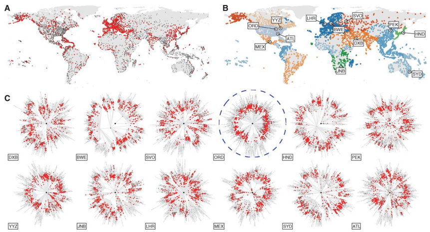
A thriving field of research



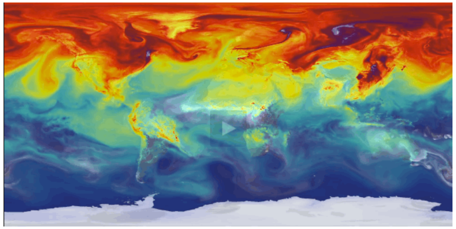
Particle physics



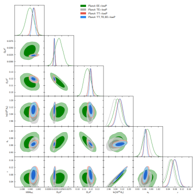
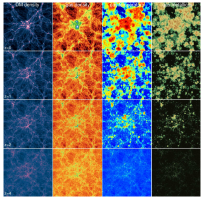
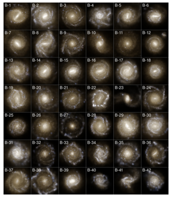
Protein folding



Epidemiology



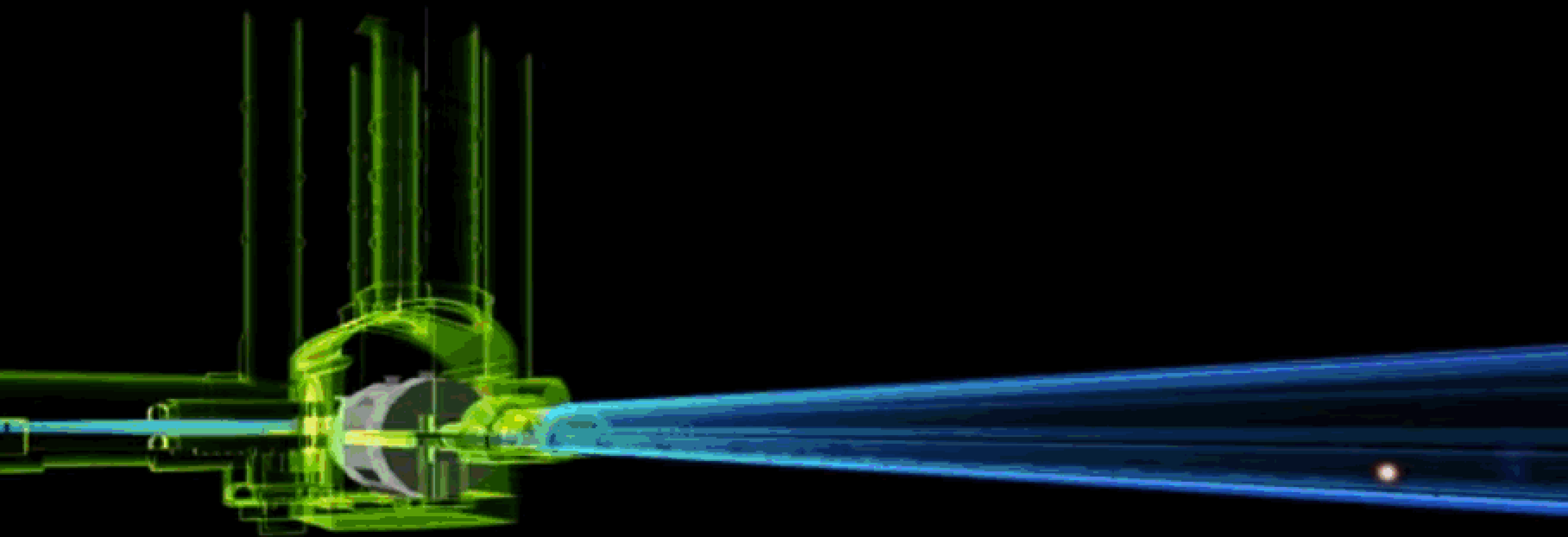
Climate science

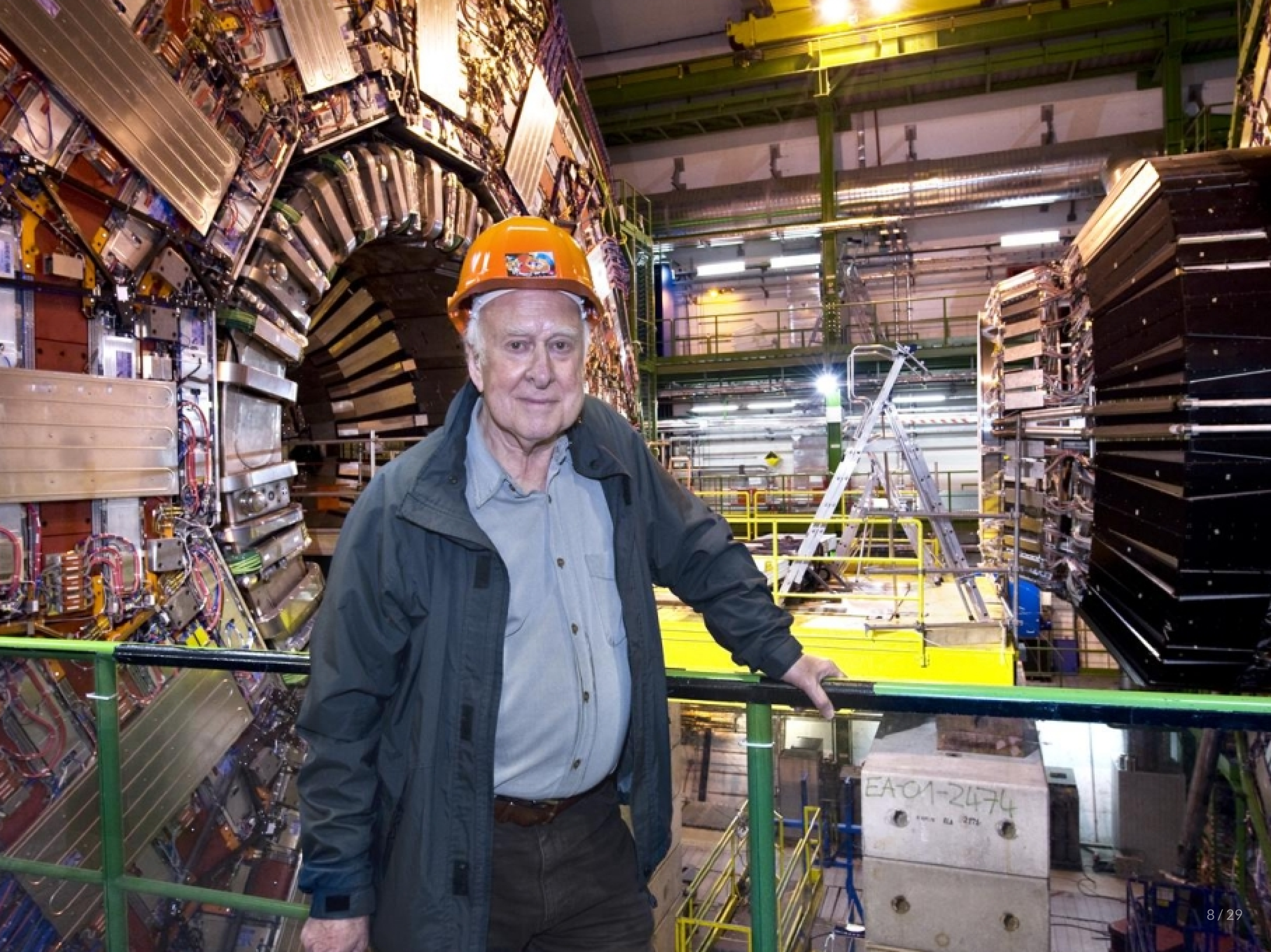




Astrophysics and cosmology

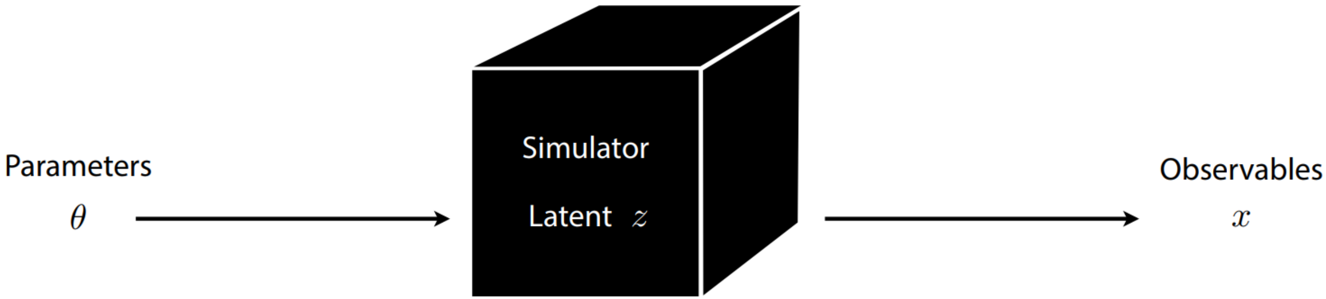
(... and many others!)

$$
\begin{gathered}
\mathcal{L}_{SM} = -\frac{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 - \frac{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 - \frac{1}{2}\partial_\nu Z^0_\mu \partial_\nu Z^0_\mu - \frac{1}{2c_w^2}M^2 Z^0_\mu Z^0_\mu - \frac{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)) - \frac{1}{2}g^2 W^+_\mu W^-_\mu W^+_\nu W^-_\nu + \frac{1}{2}g^2 W^+_\mu W^-_\nu W^+_\mu W^-_\nu + g^2c_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) - \frac{1}{2}\partial_\mu H\partial_\mu H - 2M^2\alpha_h H^2 - \partial_\mu\phi^+\partial_\mu\phi^- - \frac{1}{2}\partial_\mu\phi^0\partial_\mu\phi^0 - \\
\beta_h\left(\frac{2M^2}{g^2} + \frac{2M}{g}H + \frac{1}{2}(H^2+\phi^0\phi^0+2\phi^+\phi^-)\right) + \frac{2M^4}{g^2}\alpha_h - \\
g\alpha_h M\left(H^3 + H\phi^0\phi^0 + 2H\phi^+\phi^-\right) - \\
\frac{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)^2H^2\right) - \\
gMW^+_\mu W^-_\mu H - \frac{1}{2}g\frac{M}{c_w^2}Z^0_\mu Z^0_\mu H - \\
\frac{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) + \\
\frac{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) + \frac{1}{2}g\frac{1}{c_w}(Z^0_\mu(H\partial_\mu\phi^0 - \phi^0\partial_\mu H) + \\
M\left(\frac{1}{c_w}Z^0_\mu\partial_\mu\phi^0 + W^+_\mu\partial_\mu\phi^- + W^-_\mu\partial_\mu\phi^+\right) - ig\frac{s_w^2}{c_w}MZ^0_\mu(W^+_\mu\phi^- - W^-_\mu\phi^+) + igs_wMA_\mu(W^+_\mu\phi^- - \\
W^-_\mu\phi^+) - ig\frac{1-2c_w^2}{2c_w}Z^0_\mu(\phi^+\partial_\mu\phi^- - \phi^-\partial_\mu\phi^+) + igs_wA_\mu(\phi^+\partial_\mu\phi^- - \phi^-\partial_\mu\phi^+) - \\
\frac{1}{4}g^2W^+_\mu W^-_\mu\left(H^2+(\phi^0)^2+2\phi^+\phi^-\right) - \frac{1}{8}g^2\frac{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) - \\
\frac{1}{2}g^2\frac{s_w^2}{c_w}Z^0_\mu\phi^0(W^+_\mu\phi^- + W^-_\mu\phi^+) - \frac{1}{2}ig^2\frac{s_w^2}{c_w}Z^0_\mu H(W^+_\mu\phi^- - W^-_\mu\phi^+) + \frac{1}{2}g^2s_wA_\mu\phi^0(W^+_\mu\phi^- + \\
W^-_\mu\phi^+) + \frac{1}{2}ig^2s_wA_\mu H(W^+_\mu\phi^- - W^-_\mu\phi^+) - g^2\frac{s_w}{c_w}(2c_w^2-1)Z^0_\mu A_\mu\phi^+\phi^- - \\
g^2s_w^2A_\mu A_\mu\phi^+\phi^- + \frac{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_wA_\mu\left(-(\bar e^\lambda\gamma^\mu e^\lambda) + \frac{2}{3}(\bar u^\lambda_j\gamma^\mu u^\lambda_j) - \frac{1}{3}(\bar d^\lambda_j\gamma^\mu d^\lambda_j)\right) + \\
\frac{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)\} + \frac{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) + \\
\frac{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) + \\
\frac{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) + \\
\frac{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) - \frac{g}{2}\frac{m^\lambda_\nu}{M}H(\bar\nu^\lambda\nu^\lambda) - \\
\frac{g}{2}\frac{m^\lambda_e}{M}H(\bar e^\lambda e^\lambda) + \frac{ig}{2}\frac{m^\lambda_\nu}{M}\phi^0(\bar\nu^\lambda\gamma^5\nu^\lambda) - \frac{ig}{2}\frac{m^\lambda_e}{M}\phi^0(\bar e^\lambda\gamma^5 e^\lambda) - \frac{1}{4}\bar\nu_\lambda M^R_{\lambda\kappa}(1-\gamma_5)\hat\nu_\kappa - \\
\overline{\frac{1}{4}\bar\nu_\lambda M^R_{\lambda\kappa}(1-\gamma_5)\hat\nu_\kappa} + \frac{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) + \\
\frac{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) - \frac{g}{2}\frac{m^\lambda_u}{M}H(\bar u^\lambda_j u^\lambda_j) - \\
\frac{g}{2}\frac{m^\lambda_d}{M}H(\bar d^\lambda_j d^\lambda_j) + \frac{ig}{2}\frac{m^\lambda_u}{M}\phi^0(\bar u^\lambda_j\gamma^5u^\lambda_j) - \frac{ig}{2}\frac{m^\lambda_d}{M}\phi^0(\bar d^\lambda_j\gamma^5 d^\lambda_j) + \bar G^a\partial^2G^a + g_sf^{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 - \frac{M^2}{c_w^2})X^0 + \bar Y\partial^2 Y + igc_wW^+_\mu(\partial_\mu\bar X^0X^- - \\
\partial_\mu\bar X^+X^0) + igs_wW^+_\mu(\partial_\mu\bar YX^- - \partial_\mu\bar X^+Y) + igc_wW^-_\mu(\partial_\mu\bar X^-X^0 - \\
\partial_\mu\bar X^0X^+) + igs_wW^-_\mu(\partial_\mu\bar X^-Y - \partial_\mu\bar YX^+) + igc_wZ^0_\mu(\partial_\mu\bar X^+X^+ - \\
\partial_\mu\bar X^-X^-) + igs_wA_\mu(\partial_\mu\bar X^+X^+ - \\
\partial_\mu\bar X^-X^-) - \frac{1}{2}gM\left(\bar X^+X^+H + \bar X^-X^-H + \frac{1}{c_w^2}\bar X^0X^0H\right) + \frac{1-2c_w^2}{2c_w}igM\left(\bar X^+X^0\phi^+ - \bar X^-X^0\phi^-\right) + \\
\frac{1}{2c_w}igM\left(\bar X^0X^-\phi^+ - \bar X^0X^+\phi^-\right) + igMs_w\left(\bar X^0X^-\phi^+ - \bar X^0X^+\phi^-\right) + \\
\frac{1}{2}igM\left(\bar X^+X^+\phi^0 - \bar X^-X^-\phi^0\right) .
\end{gathered}
$$

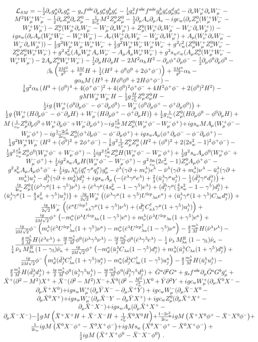



EA-01-2474

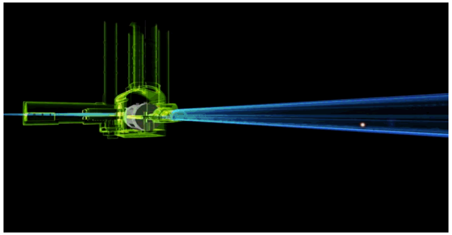
Particle physics



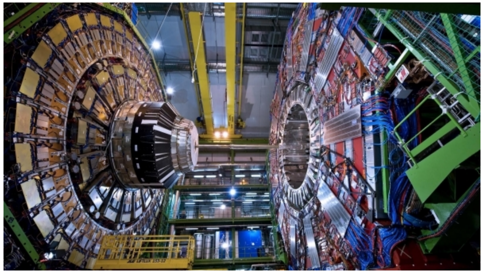
SM with parameters θ



Simulated observables x



Real observations x_{obs}



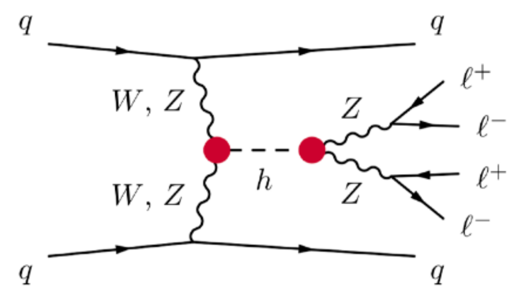
Latent variables

Parameters of interest

Parton-level momenta

Theory parameters

$$z_p \longleftarrow \theta$$



Latent variables

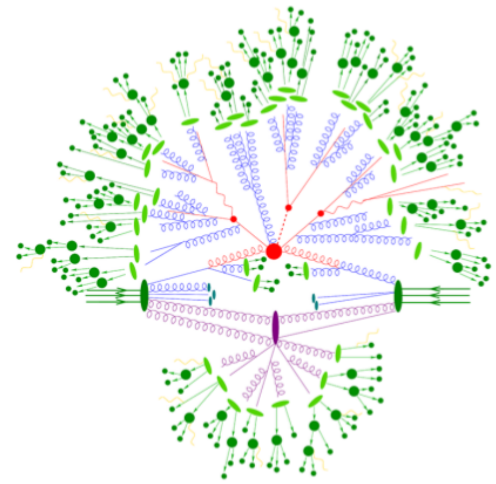
Parameters of interest

Shower splittings

Parton-level momenta

Theory parameters

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



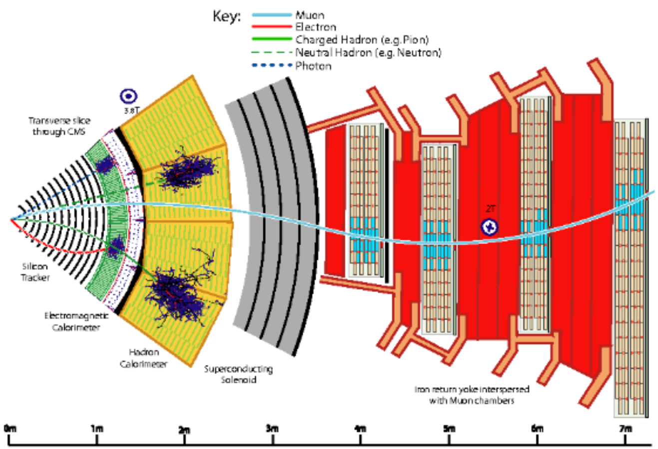
Credits: Johann Brehmer.

Latent variables

Parameters of interest

Detector interactions	Shower splittings	Parton-level momenta	Theory parameters

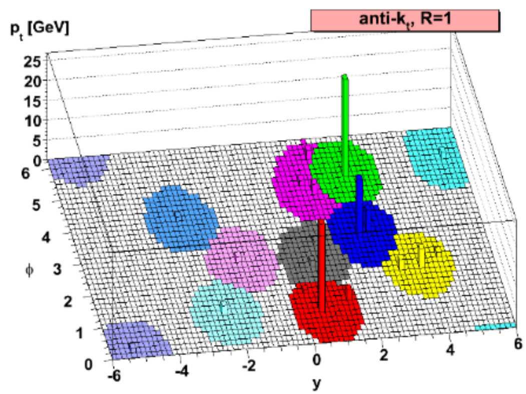
$$z_d \longleftarrow z_s \longleftarrow z_p \longleftarrow \theta$$



Features | Latent variables | Parameters of interest

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

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



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

Credits: Johann Brehmer.

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

Ingredients

Statistical inference requires the computation of key ingredients, such as

- the likelihood $p(x|\theta)$,
- the likelihood ratio $r(x|\theta_0, \theta_1) = \frac{p(x|\theta_0)}{p(x|\theta_1)}$,
- or the posterior $p(\theta|x)$.

In the simulator-based scenario, each of these ingredients can be approximated with modern machine learning techniques, even if none are tractable during training!

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
- Bayesian model comparison
- Bayesian posterior inference
- Supervised learning
- Generative adversarial networks
- Empirical Bayes with Adversarial Variational Optimization
- Optimal compression

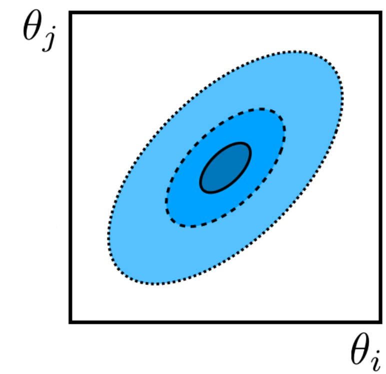
Frequentist inference

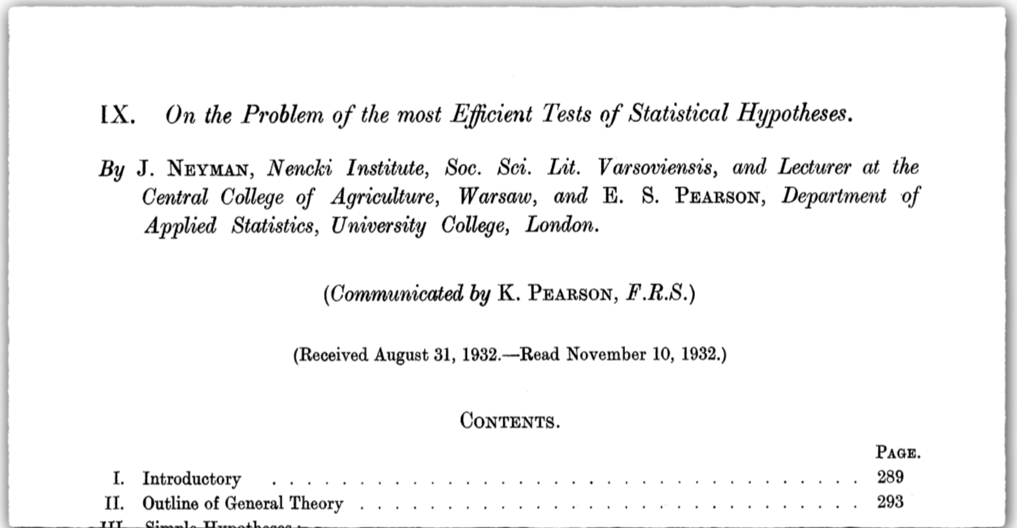
The frequentist (physicist's) way

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.*

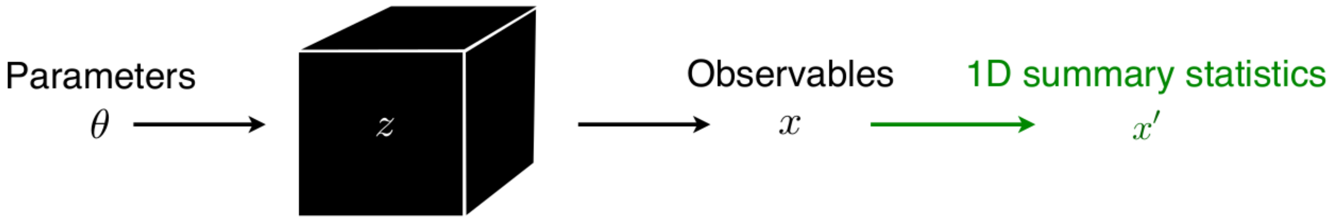
By J. Neyman, *Nencki Institute, Soc. Sci. Lit. Varsoviensis, and Lecturer at the Central College of Agriculture, Warsaw, and* E. S. Pearson, *Department of Applied Statistics, University College, London.*

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

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

Contents.

		Page.
I.	Introductory	289
II.	Outline of General Theory	293



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

Then, approximate the likelihood $p(x|\theta)$ with the surrogate $\hat{p}(x|\theta) = p(x'|\theta)$.

From this it comes

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

Wilks theorem

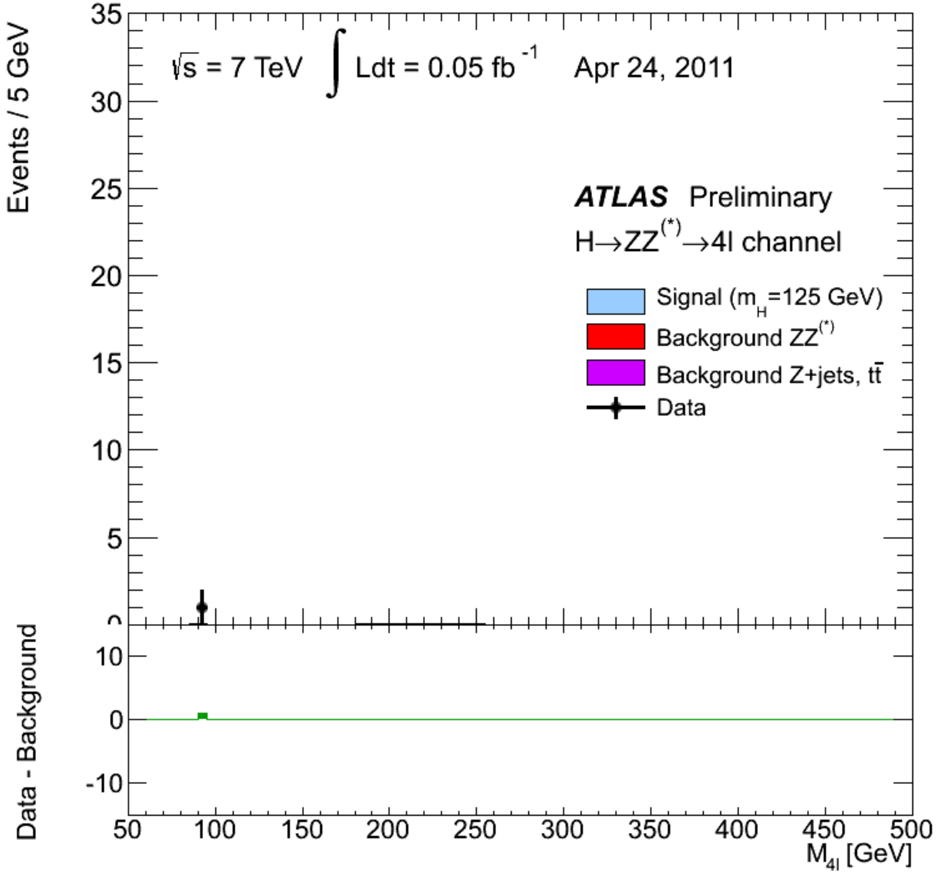
Consider the test statistic

$$q(\theta) = -2 \sum_x \log \frac{p(x|\theta)}{p(x|\hat{\theta})} = -2 \sum_x \log r(x|\theta, \hat{\theta})$$

for a fixed number N of observations $\{x\}$ and where $\hat{\theta}$ is the maximum likelihood estimator.

When $N \to \infty, q(\theta) \sim \chi_2$.

Therefore (and provided the assumptions apply!), an observed value $q_{\mathrm{obs}}(\theta)$ translates directly to a p-value that measures the confidence with which θ can be excluded.

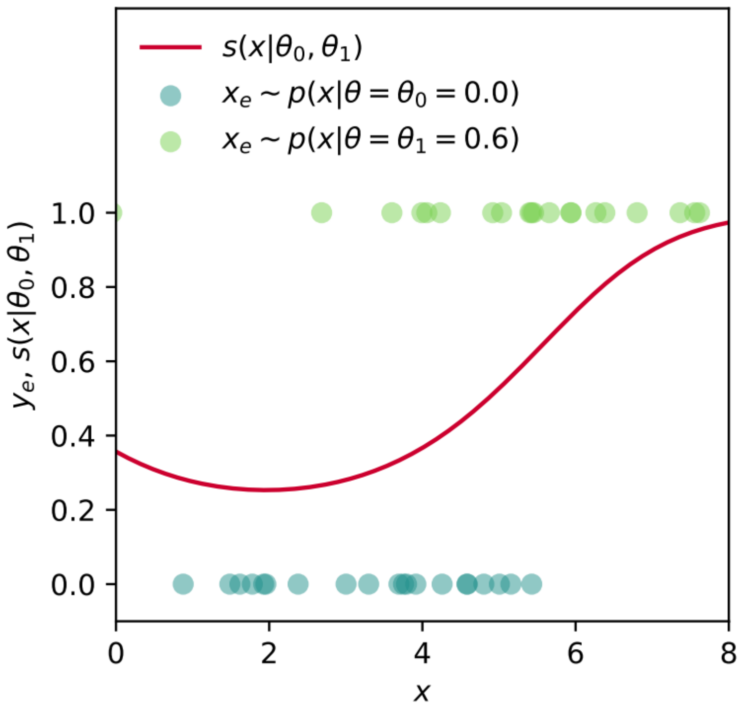


Discovery of the Higgs boson at 5-σ

CARL

Supervised learning provides a way to automatically construct s:

- Let us consider a neural network classifier $\hat{s}$ tasked to distinguish $x_i \sim p(x|\theta_0)$ labelled $y_i = 0$ from $x_i \sim p(x|\theta_1)$ labelled $y_i = 1$.
- Train $\hat{s}$ by minimizing the cross-entropy loss.



Cranmer, Pavez and Louppe, 2015 [arXiv:1506.02169].

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

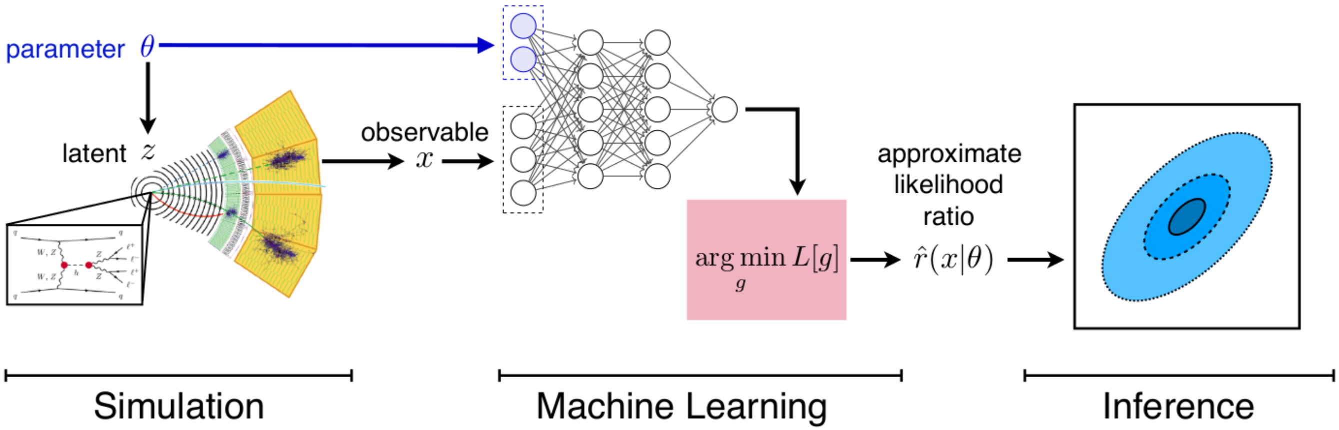
$$\hat{s}(x) \approx s^*(x) = \frac{p(x|\theta_1)}{p(x|\theta_0) + p(x|\theta_1)}.$$

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.

Cranmer, Pavez and Louppe, 2015 [arXiv:1506.02169].



To avoid retraining a classifier $\hat{s}$ for every (θ_0, θ_1) pair, fix θ_1 to θ_{ref} and train a single parameterized classifier $\hat{s}(x|\theta_0, \theta_{\text{ref}})$ where θ_0 is also given as input.

Therefore, we have

$$\hat{r}(x|\theta_0, \theta_{\text{ref}}) = \frac{1 - \hat{s}(x|\theta_0, \theta_{\text{ref}})}{\hat{s}(x|\theta_0, \theta_{\text{ref}})}$$

such that for any (θ_0, θ_1),

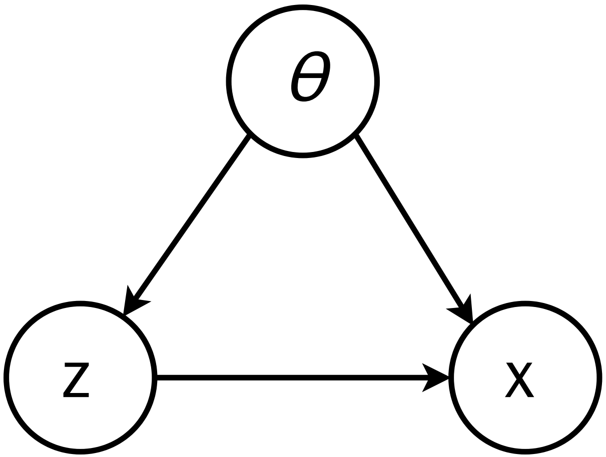
$$r(x|\theta_0, \theta_1) \approx \frac{\hat{r}(x|\theta_0, \theta_{\text{ref}})}{\hat{r}(x|\theta_1, \theta_{\text{ref}})}.$$

Cranmer, Pavez and Louppe, 2015 [arXiv:1506.02169].

Bayesian inference

Bayesian inference = computing the posterior

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



Doubly intractable in the likelihood-free scenario:

- Cannot evaluate the likelihood $p(x|\theta) = \int p(x, z|\theta)dz$.
- Cannot evaluate the evidence $p(x) = \int p(x|\theta)p(\theta)d\theta$.

Amortizing Bayes

The Bayes rule can be rewritten as

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} = r(x|\theta)p(\theta) \approx \hat{r}(x|\theta)p(\theta),$$

where $r(x|\theta) = \frac{p(x|\theta)}{p(x)}$ is the likelihood-to-evidence ratio.

Amortizing Bayes

The Bayes rule can be rewritten as

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} = r(x|\theta)p(\theta) \approx \hat{r}(x|\theta)p(\theta),$$

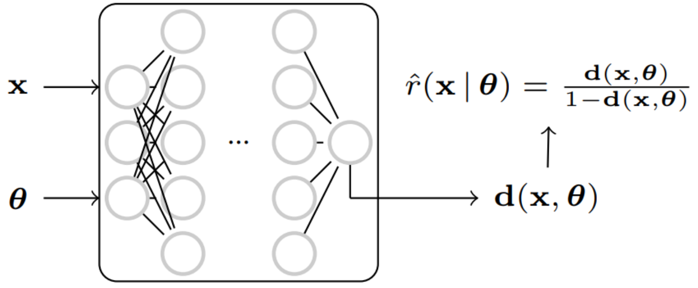
where $r(x|\theta) = \frac{p(x|\theta)}{p(x)}$ is the likelihood-to-evidence ratio.

As before, the likelihood-to-evidence ratio can be approximated e.g. from a neural network classifier trained to distinguish $x \sim p(x|\theta)$ from $x \sim p(x)$, hence enabling direct and amortized posterior evaluation.

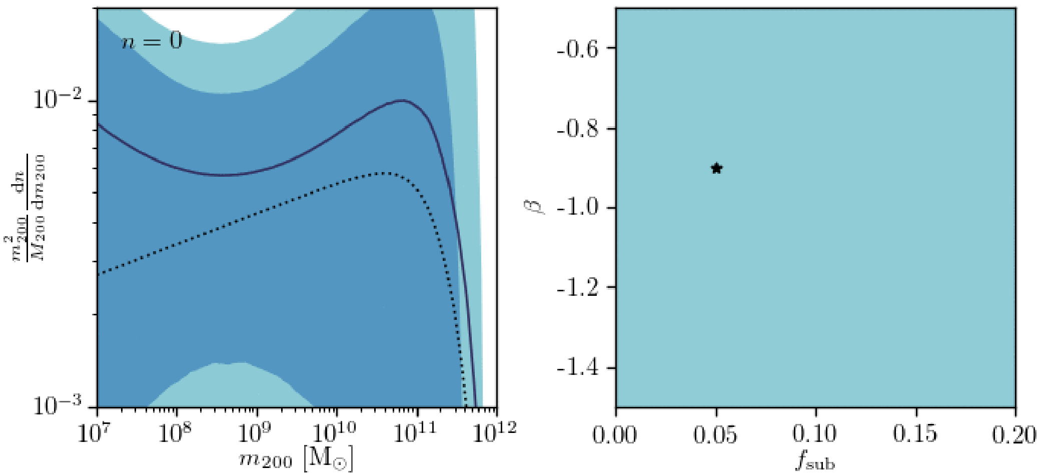
Algorithm 1 Optimization of $\mathbf{d}(\mathbf{x}, \boldsymbol{\theta})$.

Inputs:	Criterion ℓ (e.g., BCE)
	Implicit generative model $p(\mathbf{x} \mid \boldsymbol{\theta})$
	Prior $p(\boldsymbol{\theta})$
Outputs:	Parameterized classifier $\mathbf{d}_\phi(\mathbf{x}, \boldsymbol{\theta})$
Hyperparameters:	Batch-size M

while not converged **do**
    Sample $\boldsymbol{\theta} \leftarrow \{\boldsymbol{\theta}_m \sim p(\boldsymbol{\theta})\}_{m=1}^M$
    Sample $\boldsymbol{\theta}' \leftarrow \{\boldsymbol{\theta}'_m \sim p(\boldsymbol{\theta})\}_{m=1}^M$
    Simulate $\mathbf{x} \leftarrow \{\mathbf{x}_m \sim p(\mathbf{x} \mid \boldsymbol{\theta}_m)\}_{m=1}^M$
    $\mathcal{L} \leftarrow \ell(\mathbf{d}_\phi(\mathbf{x}, \boldsymbol{\theta}), 1) + \ell(\mathbf{d}_\phi(\mathbf{x}, \boldsymbol{\theta}'), 0)$
    $\phi \leftarrow \text{OPTIMIZER}(\phi, \nabla_\phi \mathcal{L})$
end while
return $\mathbf{d}_\phi$

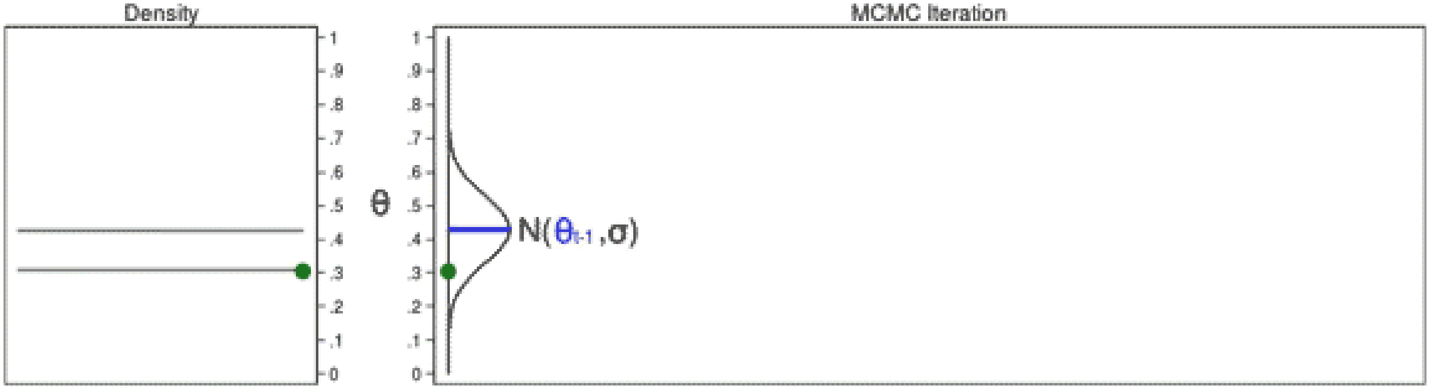


Bayesian inference of dark matter subhalo population parameters



Brehmer, Mishra-Sharma, Hermans, Louppe, and Cranmer, 2019 [arXiv:1909.02005].

MCMC posterior sampling



Step 1: $r(\theta_{new}, \theta_{t-1}) = \dfrac{\text{Posterior}(\theta_{new})}{\text{Posterior}(\theta_{t-1})} = \dfrac{\text{Beta}(1,1,0.306) \times \text{Binomial}(10,4,0.306)}{\text{Beta}(1,1,0.429) \times \text{Binomial}(10,4,0.429)} = 0.834$

Step 2: Acceptance probability $\alpha(\theta_{new}, \theta_{t-1}) = \min\{r(\theta_{new}, \theta_{t-1}), 1\} = \min\{0.834, 1\} = 0.834$

Step 3: Draw $u \sim \text{Uniform}(0,1) = 0.617$

Step 4: If $u < \alpha(\theta_{new}, \theta_{t-1}) \rightarrow$ If $0.617 < 0.834$ Then $\theta_t = \theta_{new} = 0.306$
Otherwise $\theta_t = \theta_{t-1} = 0.429$

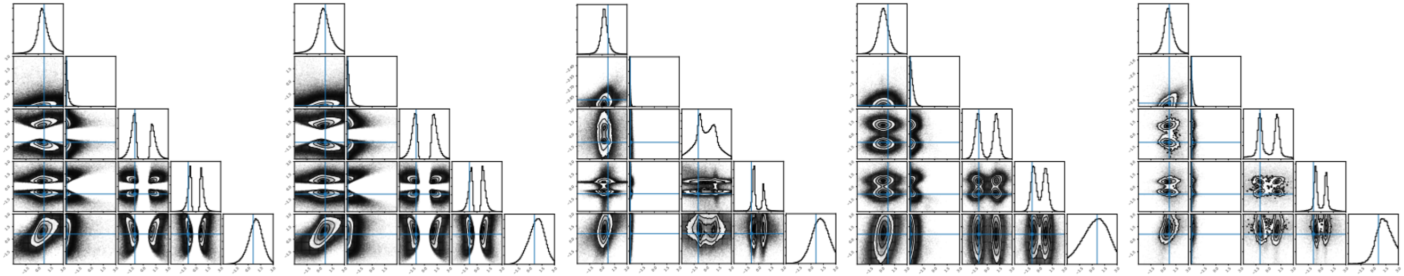
Likelihood-free MCMC

MCMC samplers require the evaluation of the posterior ratios:

$$\begin{aligned}
\frac{p(\theta_{\text{new}}|x)}{p(\theta_{t-1}|x)} &= \frac{p(x|\theta_{\text{new}})p(\theta_{\text{new}})/p(x)}{p(x|\theta_{t-1})p(\theta_{t-1})/p(x)} \\
&= \frac{p(x|\theta_{\text{new}})p(\theta_{\text{new}})}{p(x|\theta_{t-1})p(\theta_{t-1})} \\
&= r(x|\theta_{\text{new}}, \theta_{t-1})\frac{p(\theta_{\text{new}})}{p(\theta_{t-1})}
\end{aligned}$$

Again, MCMC samplers can be made likelihood-free by plugging a learned approximation $\hat{r}(x|\theta_{\text{new}}, \theta_{t-1})$ of the likelihood ratio.

Hermans, Begy and Louppe, 2019 [arXiv:1903.04057].



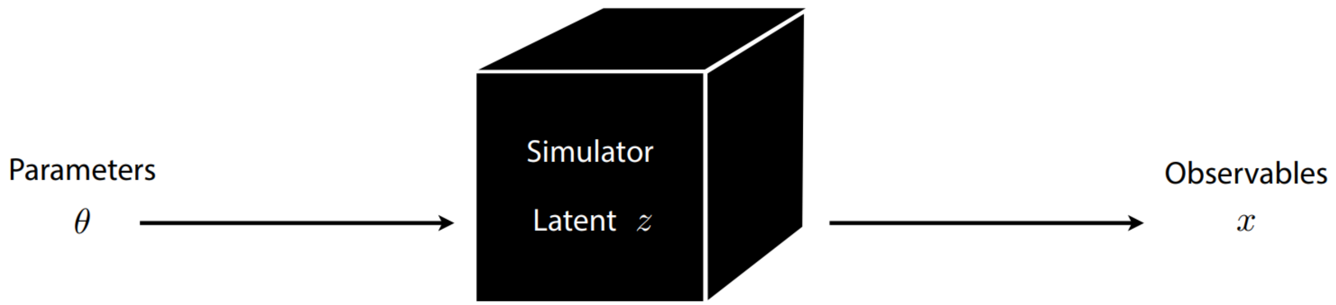
(a) MCMC groundtruth (b) AALR-MCMC (c) SNPE-A (d) SNPE-B (e) APT

Figure 3: Posteriors from the tractable benchmark. The experiments are repeated 25 times and the approximate posteriors are subsampled from those runs. AALR-MCMC shares the same structure with the MCMC truth, demonstrating its accuracy. Some runs of the other methods were not consistent, contributing to the variance observed in Table 2.

Hermans, Begy and Louppe, 2019 [arXiv:1903.04057].

Summary

- Much of modern science is based on "likelihood-free" simulations.
- The likelihood-ratio is central to many statistical inference procedures, regardless of your religion.
- Supervised learning enables likelihood-ratio estimation.



Collaborators


Kyle Cranmer


Juan Pavez


Johann Brehmer


Joeri Hermans


Siddarth Mishra-Sharma

References

- Brehmer, J., Mishra-Sharma, S., Hermans, J., Louppe, G., Cranmer, K. (2019). Mining for Dark Matter Substructure: Inferring subhalo population properties from strong lenses with machine learning. arXiv preprint arXiv 1909.02005.
- Hermans, J., Begy, V., & Louppe, G. (2019). Likelihood-free MCMC with Approximate Likelihood Ratios. arXiv preprint arXiv:1903.04057.
- Baydin, A. G., Shao, L., Bhimji, W., Heinrich, L., Meadows, L., Liu, J., ... & Ma, M. (2019). Etalumis: Bringing Probabilistic Programming to Scientific Simulators at Scale. arXiv preprint arXiv:1907.03382.
- 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.

The end.