

Faculty of Physics

Testing machine learning against finite size scaling using MAFs

Simran Singh *University of Bielefeld*

work done as part of **A01** *project in* **CRC Tr-211** *between Bielefeld (F. Karsch, C. Schmidt & S. Singh) and Frankfurt (O. Phillipsen, R. Kaiser, J.P. Klinger)*

ML meets LFT @ Swansea University July 24-26, 2024

Outline

- *I. Motiv*a*tion N*a*ture of the chir*a*l tr*a*nsition*
- *II. Nf = 5 project using MAFs* a*nd HISQ (old)*
- *III.* MADE and MAF
- *IV.* (new) Nf = 5 using Stout smeared data
- *V. Results on density estim*a*tion*
- *VI. Future directions*

N_f and the chiral transition

- Nature of the QCD chiral phase transition is (strictly speaking) inconclusive
- Cannot simulate $m_f = 0$ on lattice can only extrapolate from finite mass simulations
- Proposal for studying the critical surface that separates first-order regions from crossover as function of degenerate N_f quarks by [F. Cuteri et.al.*, JHEP* 11 (2021)]
- Using the argument that the surface terminates in a tri-critical line, conclusions for the order of the chiral transition can be drawn.

Z2 boundary for Nf=5 HISQ

- The analysis of the above kind requires LOTS of lattice simulations for EACH $N_{\hspace{-0.8pt}f}$, varying lattice volumes and spacings need to be studied **1. Introduction**
- In [M. Neumann et.al., PoS LATTICE2022 (2023)], the authors proposed using Machine Learning techniques for learning probability densities p $(\bar{\psi}\psi,S)$ conditioned on N_{σ} , m_l , β Cuteri *et al.* finds that in the continuum limit there is no first order transition for light quark masses
- Learning such a density correctly allows interpolation in the dimensions of the conditional inputs - avoiding some expensive lattice simulations

Z2 boundary for Nf=5 HISQ histogram of the action and the action and the observable we want to reweight. Moreover, the action histograms
The action histograms of the action histograms of the action histograms of the action histograms of the action **Figure 7:** Phase diagram of 5-flavor QCD on lattices with fixed temporal Ω section we will discuss Ω a ML based approach to lo-

imann et al., PoS LATTICF2022 (2023) \vert Neumann M (2

M. Neumann et.al.*, PoS* LATTICE2022 (2023) | Neumann M (2023) PhD Thesis Universität Bielefeld

gap between them. In the main them. In the main

5

• First step (Will discuss): Density estimation followed by β , m_l , N_σ extrapolation using *Masked Autoregressive flows* $\overline{}$ method can be extended to reweight a probability distribution of any observable by $\overline{}$ **ig** Masked Au

• Second step (Won't discuss): Classification of densities projected as "images" via ransformers to nail down m_i for $N_c = 5$ vision transformers to nail down $m_{l,c}$ for $N_f = 5$

Density estimation using MADE \overline{O} and \overline{O} . The case of \overline{O} Universite de Sherbrooke, Canada ´ \mathcal{L} A $\sum_{i=1}^n$

MADE: Masked Autoencoder for Distribution Estimation Universite de Sherbrooke, Canada ´ $\mathbb{E}[\mathbf{U}(\mathbf{U})] = \mathbf{U}[\mathbf{U}(\mathbf{U})]$ A brief description of the basic autoencoder, on which this

2. Autoencoders

Mathieu Germain Karol Gregor Iain Murray Hugo Larochelle **COM** Machine Lea ked Autoencoder for Distribution Estimation $\sum_{n=0}^{\infty}$ DoS of 32nd International Conference Karol Gregor Iain Murray Hugo Larochelle **Hugo.** Hugo Larochelle p ation p p **on Machine Learning, France, 2015**

Mathieu Germain MATHIEU.GERMAIN2@USHERBROOKE.CA

valid probabilities. Specifically, we'd like to be able to be able to be able to write to write to write to wri

examples *{*x(*t*) *}T* $\frac{1}{2}$

Made: Ma
Estimation Estimation Estimation Estimation Estimation Estimation Estimation Estimation Estimation Estimation

Universite de Sherbrooke, Canada ´

- C_{α} de U_{α} Example Gregorian G • Goal : Learn a probability density from examples of data $(\vec{x}, \vec{y}) \rightarrow p(\vec{x} | \vec{y})$ input dimension *x^d* belongs in *{*0*,* 1*}*. The motivation is am examples of data $(\vec{x} \ \vec{v}) \rightarrow p(\vec{x} | \vec{v})$ it is into the product of its nested condition \mathcal{L} ensity from examples of data $(x,$
- \bullet How : Interpret the outputs of an autoencoder as valid prof University of Edinburgh, United Kingdom University of Edinburgh, United Kingdom erpret the outputs of an autoencoder as valid probabilities Universite de Sherbrooke, Canada ´ neural network models to estimate a distribution • How : Interpret the outputs of an autoencoder as valid probabilities atto use interpret the extents of an **Priow**. merpret the duputs of an autocheduci as vaila p mated the quipute of an autoencoder as valid probabilities *as* vand probabilities

Google Deep Minds

- Hugo Larochelle HUGO.LAROCHELLE@USHERBROOKE.CA University of $\mathsf{L}\mathsf{d}\mathsf{C}\mathsf{H}$ ϵ Each output as senditional probability and product as ioint Lach bulput as anal probability and pro onal probability and pro \mathcal{L} Domingos, 2011; Dinh et al., 2014), data (e.g. speech), data (e.g. speech as joint probability. The very other from a set of examples. We introduce a simple \mathbf{r} pability and product as joint probability yields powerful generative models. Our method • Each output as conditional probability and product as joint probability representation h(x) of its input x such that, from it, we can expect the such that, from it, we can expect the $\frac{1}{2}$ *d*=1 *a k as conditional probability and product as joint probability* adpendes conditional probability and *duct as joint probability* matrix product, we have:
- \mathcal{M} auce **masks** on nidden layer units to impos Abstract et al., 2009), denoising or missing input imputation (Poon asks on hidden laver units neural network models to estimate a distribution of the state and distribution of the state and distribution of a Dominista Dinh et al., 2014), proporty ipose **autoregressive** property ddon lavor unite to impo agen layer anno to mipo challenge for machine learning. In essence, the curse of the curse itoregressive property masks the autoencoder's parameters to respect er units to impose **autoregressive** property **• Introduce masks** on hidden layer units to impose **autoregressive** property \cdots \cdots \cdots \cdots *its to impose autoregrees* there must be no computational path between output unit *x*ˆ*^d* and any of the input units *xd,...,xD*. In other words, *K K*

Masked Autoregressive Flows

- **Autoregressive property** from conditionals $p(x_1, x_2 \ldots x_D) = p(x_N | x_1, \ldots x_{N-1}) p(x_{N-1} | x_1, \ldots x_{N-2}) \ldots p(x_1)$
- Each conditional as a single Gaussian : $p(x_i | \vec{x}_{1:i-1}) = \mathcal{N}(x_i | \mu_i, (\exp(\alpha_i))^2)$ with $\mu_i = f_{\mu_i}(\vec{x}_{1:i-1})$ and $\alpha_i = f_{\alpha_i}(\vec{x}_{1:i-1})$ ⃗
- Data generated via : $x_i = u_i \exp(\alpha_i) + \mu_i$ with $u_i \sim \mathcal{N}(0,1)$
- A flow is then constructed by MADE blocks in a chain

7

The goal : Test the procedure for different data

- Goal : To reproduce the Z2 critical boundary via ML for [F. Cuteri et.al., *JHEP* 11 (2021)]
- Un-improved staggered quarks $N_f = 5$, $N_\tau = 4$ with $N_\sigma \in \{8, 12, 16\}$ and *m*_{*i*} ∈ {0.075, 0.080, 0.085, 0.090}
- Trained only on $N_{\sigma} \in \{8, 16\}$, total training data ~3.4 million values for $(\bar{\psi}\psi, S)$

Results : $\langle \bar{\psi}\psi \rangle$ for $N_{\sigma} = 8$

- Training done by removing **all** $N_{\sigma} = 12$ data
- Quantity obtained : $p(\bar{\psi}\psi, S | N_{\sigma}, m_l, \beta)$
- Results for 100K evaluations of the model

MAF prediction for the β interpolation on training set

9

Results for $\langle \bar{\psi}\psi \rangle$ for $N_{\sigma} = 16$

MAF prediction for the β interpolation

Results : ⟨*ψ*¯ *ψ*⟩

MAF prediction for $N_{\sigma} = 12$ (genuine prediction !)

11

Results : $\chi_{\bar{y}y}$ for $8^3 \times 4$

- With $p(\bar{\psi}\psi, S | N_{\sigma}, m_l, \beta)$ we are free to compute higher moments !
- We see scaling of peak height, width, location from ML prediction

Results for $\chi_{\bar{\psi}\psi}$ for 16^3 , $12^3 \times 4$

Results for $p(\bar{\psi}\psi, S)$ for some N_{σ}, m_{l}, β

MAF Inference vs standard finite size scaling from lattice studies

Ongoing steps

- Currently the model doesn't (want to) train on the $N_{\sigma} = 12$ data
- Only running on TensorFlow for CPUs required package not compatible with current TF version & Model doesn't compile on new TF version
- Expand the conditionals to N_{τ} and N_f
- Explore in the direction of a statement made in [G. Papamakarios et. al., [1705.07057](https://doi.org/10.48550/arXiv.1705.07057)]

"… accurate densities do not necessarily imply good performance in other tasks, such as in data generation … Choice of method should be informed by whether the application at hand calls for accurate densities, latent space inference or high quality samples "

A possible direction ? Here, Udtavia is the temporal Wilson line between the temporal Wilson line between the temporal Wilson line be
Here, Udtavia is the temporal Wilson line between the temporal Wilson line between the temporal Wilson line be PHYSICAL REVIEW LETTERS 130, 231902 (2023)

- Learning probability densities for correlators typically needed in spectral function reconstruction …ς ασπριείσετες τοι σοπισιατοις εχρισαπ, ποσάσε
uction $f(x)$. Generalization at $f(x)$ finite renormalization at $f(x)$ finite renormalization at $f(x)$ ander than the smaller $\sum p \cdot c \cdot a$
- Can we make G_E "continuous" in T to get a better re-constructed spectral function ? θ continuum θ in τ to cot election θ constant $\frac{E}{E}$ continuous in z to get a better re-constru \mathbf{S} inverting \mathbf{S} $p = 3$ showled be respected by $p = 3$.

$$
G_E(\tau,T)=\int_0^\infty\frac{\mathrm{d}\omega}{\pi}\rho_E(\omega,T)\frac{\cosh[\omega\tau-\omega/(2T))]}{\sinh[\omega/(2T)]},
$$

L. Altenkort et. al., PHYSICAL REVIEW LETTERS 130, 231902 (2023)

suppressed for ffiffiffiffiffiffiffi