LEARNING NEW PHYSICS FROM A MACHINE

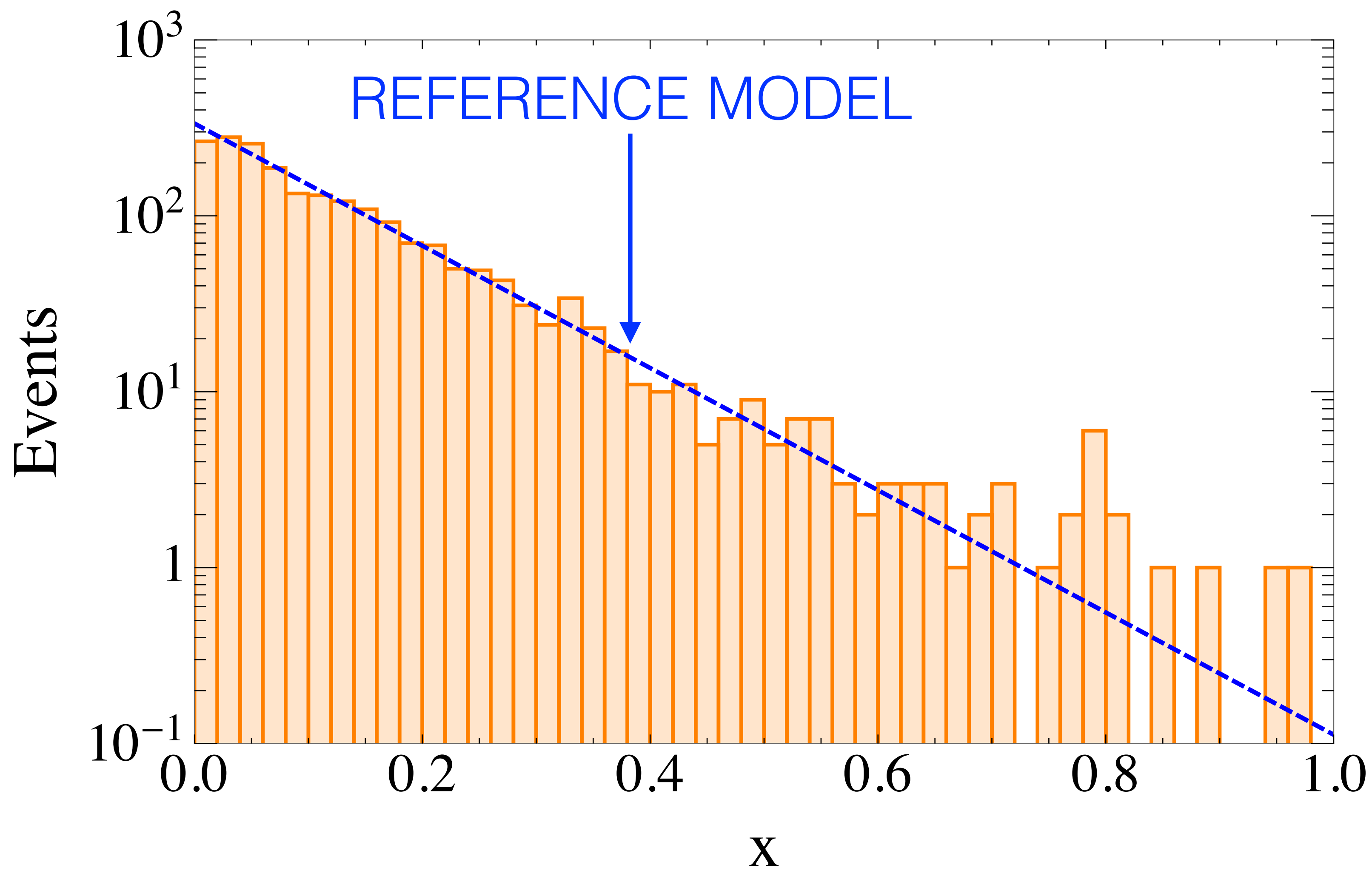
Raffaele Tito D'Agnolo - SLAC

CIPANP 2018

Raffaele Tito D'Agnolo and Andrea Wulzer

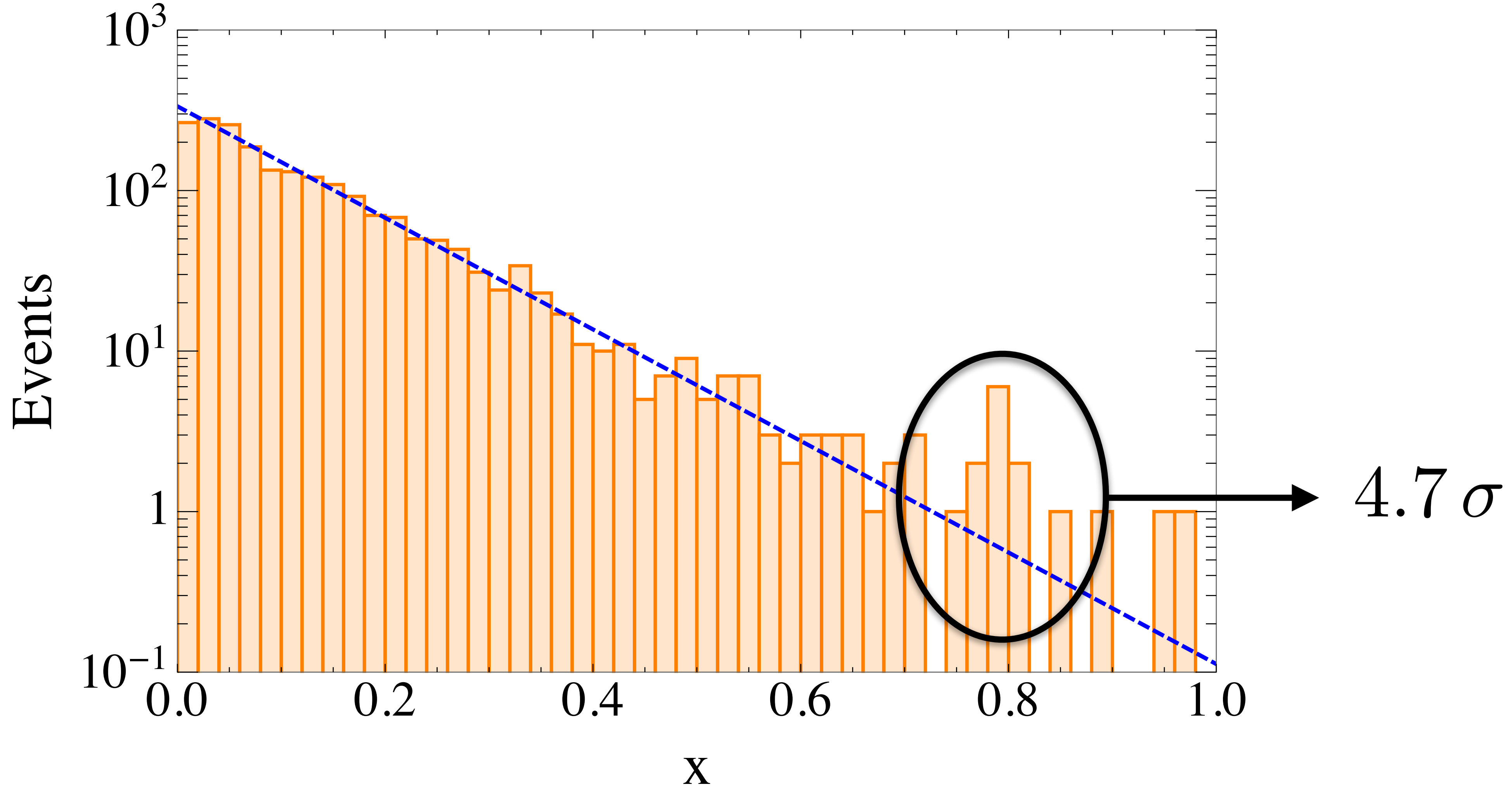
To Appear (next week)

THE PROBLEM



$$\chi^2 = 47 \quad N_{\text{bins}} = 50 \quad p - \text{value} < 1\sigma$$

THE PROBLEM



$$t_{\rm id}(\mathcal{D}) = 2\,\log\left[\frac{e^{-N(\rm NP)}}{e^{-N(\rm R)}}\prod_{x\in\mathcal{D}}\frac{n(x|\rm NP)}{n(x|\rm R)}\right]$$

WHAT IS A NEURAL NETWORK?

SET OF FUNCTIONS

+

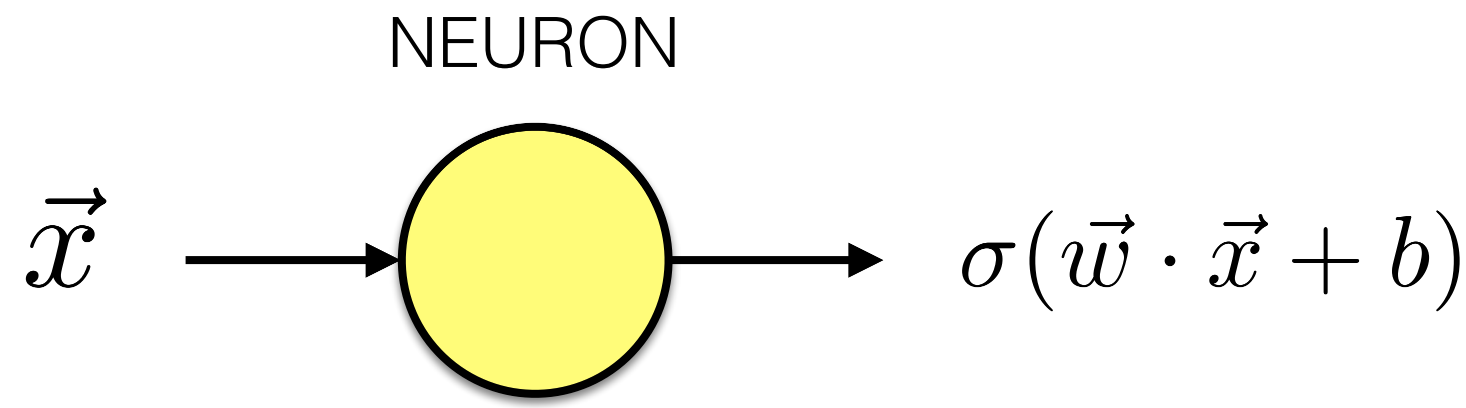
FITTING ALGORITHM

WHAT IS A NEURAL NETWORK?

SET OF FUNCTIONS

$$f_{w_1}^{(1)}\left(f_{w_2}^{(2)}\left(f_{w_3}^{(3)}(\ldots)\right)\right)$$

BUILDING BLOCKS



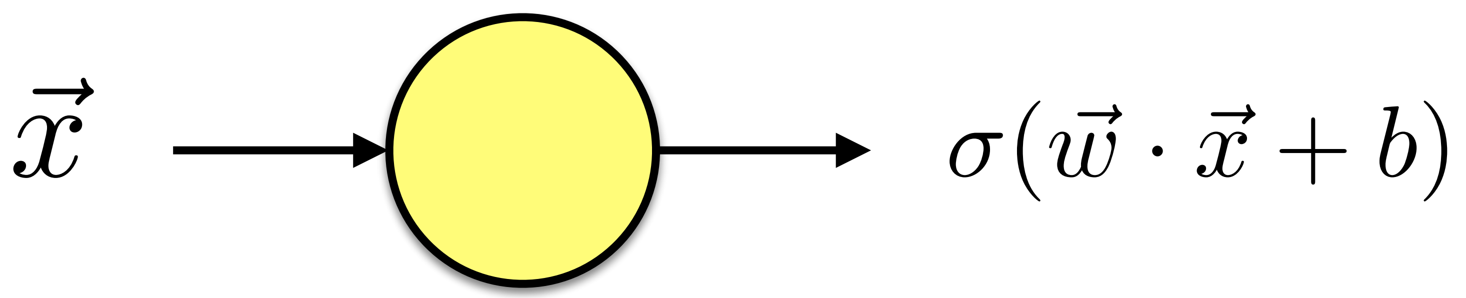
1. LINEAR TRANSFORMATION $z = \vec{w} \cdot \vec{x} + b$

   FREE PARAMETERS

2. NON-LINEAR TRANSFORMATION $\sigma(z)$ FIXED

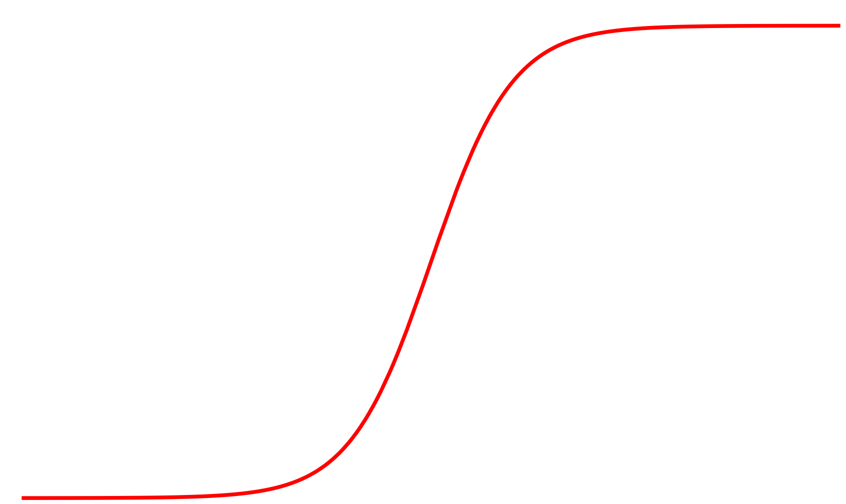
BUILDING BLOCKS

NEURON



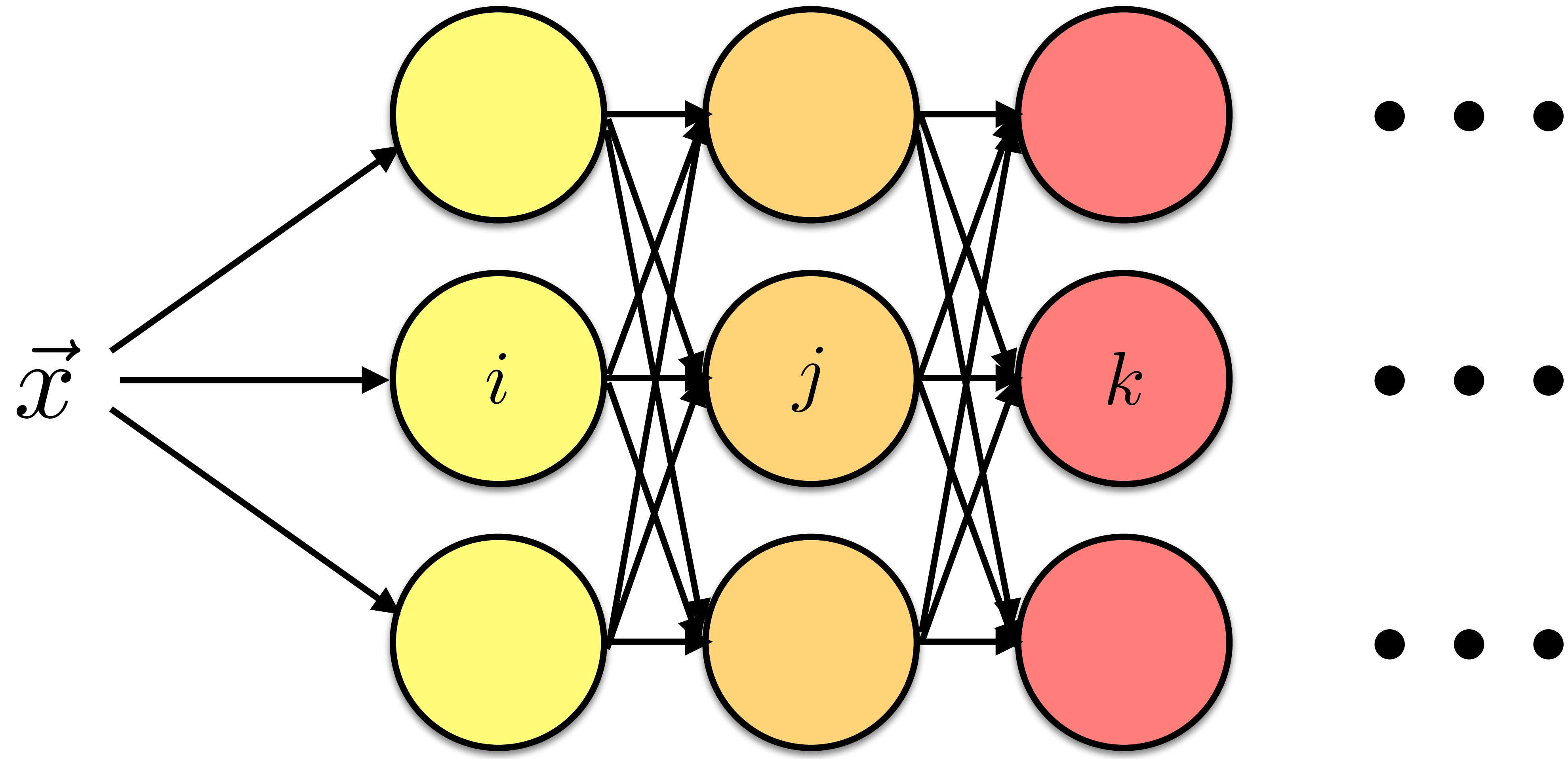
2. NON-LINEAR TRANSFORMATION

$$\sigma(z) = \begin{cases} \tanh(z) \\ \text{ReLU} \\ \frac{1}{1+e^{-z}} \\ \ldots \end{cases}$$



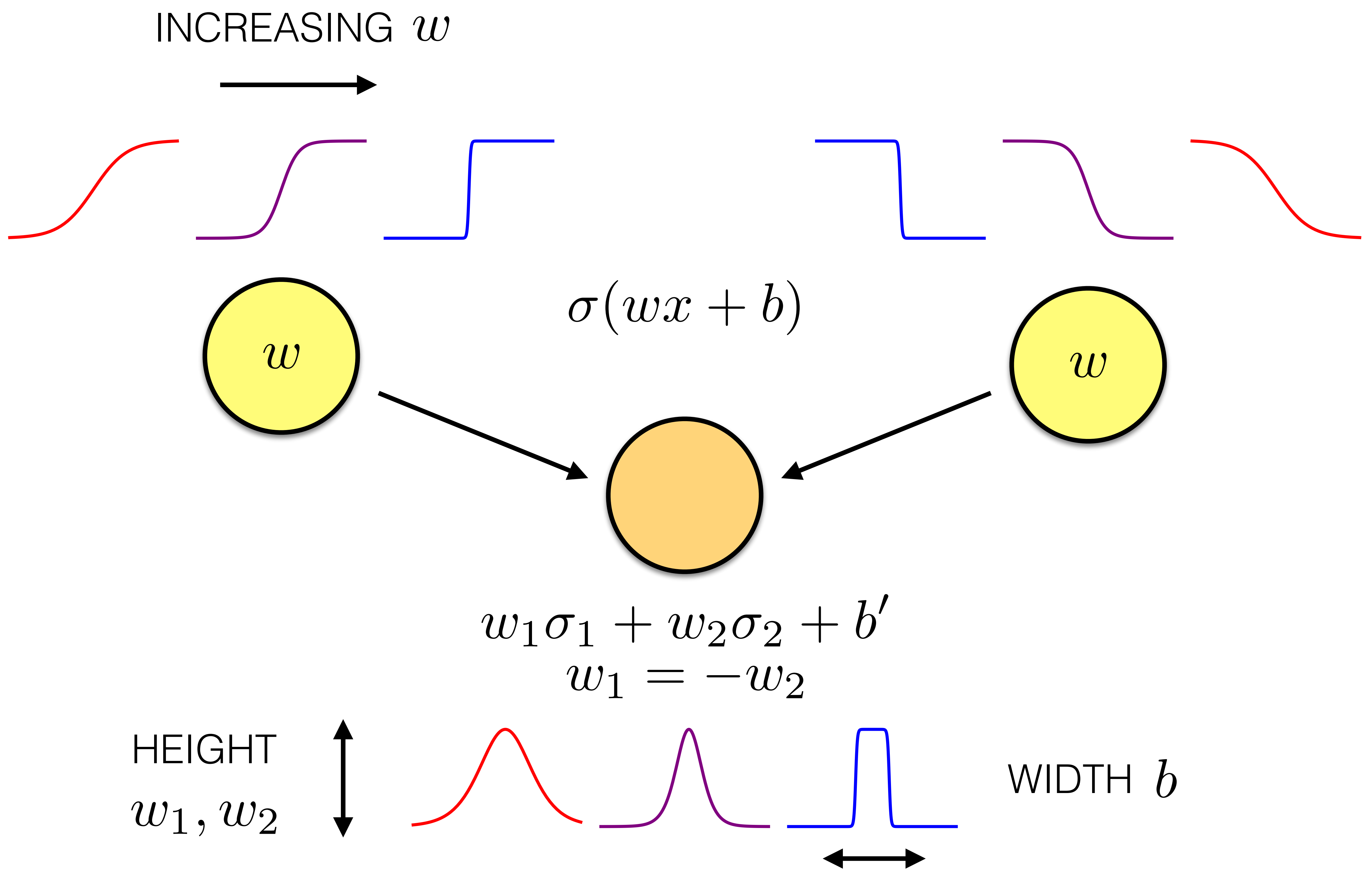
THE NETWORK

FEEDFORWARD, FULLY CONNECTED



$$\sigma\left(\sum_{k=1}^{3} w_{jk}\sigma_k\left(\sum_{i=1}^{3} w_{ki}\sigma_i\left(\sum_{l=1}^{d} w_{il}x_l + b_i\right) + b_k\right) + b_j\right)$$

UNIVERSAL APPROXIMANTS



FITTING ALGORITHM

LOSS FUNCTION

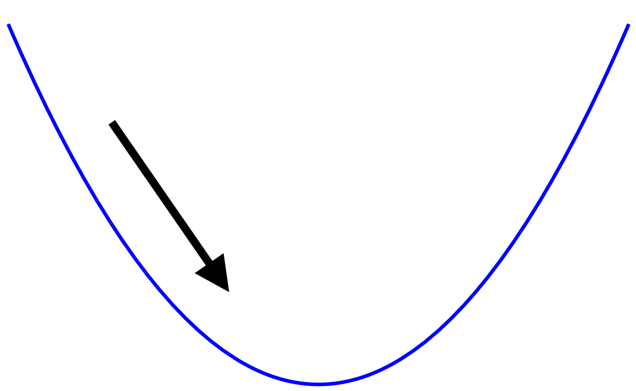
$$L = \frac{1}{N_c}\sum_{i=1}^{N_c}[1 - f_{NN}(\vec{x}_i, \mathbf{w}, \mathbf{b})]^2 + \frac{1}{N_d}\sum_{j=1}^{N_d}[f_{NN}(\vec{x}_j, \mathbf{w}, \mathbf{b})]^2$$

TRAINING

$$w_{t+1} \to w_t - \epsilon \partial_w \hat{L}$$

$\hat{L}$ SUBSET OF THE SAMPLE

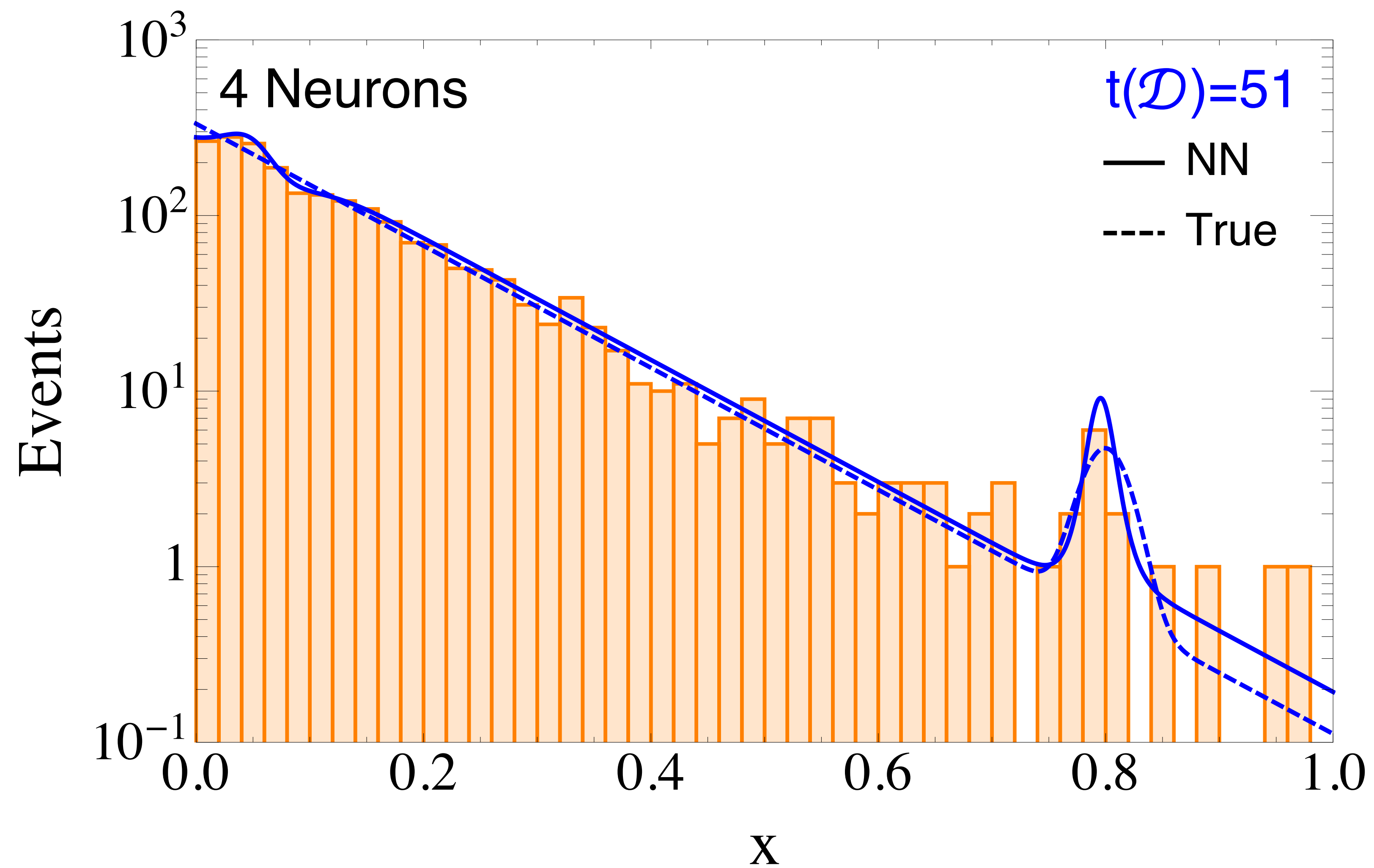
ϵ LEARNING RATE



LEARNING NEW PHYSICS

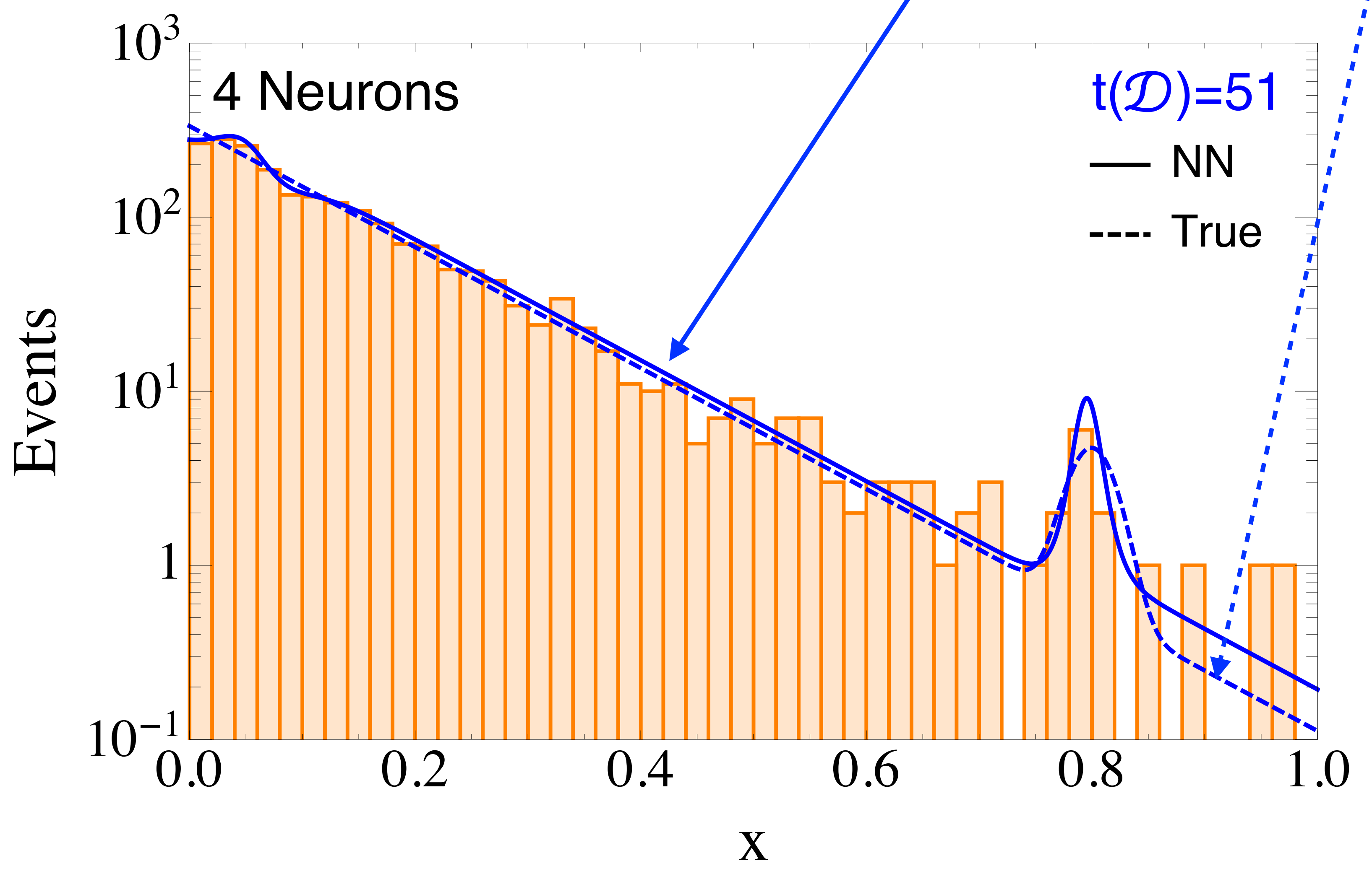
A SIMPLE STRATEGY

BINNED HISTOGRAM → SMOOTH APPROXIMANT



A SIMPLE STRATEGY

1. LEARN THE DATA DISTRIBUTION $n(x|\widehat{\mathbf{w}}) \approx n(x|\mathrm{T})$



A SIMPLE STRATEGY

1. LEARN THE DATA DISTRIBUTION $n(x|\widehat{\mathbf{w}}) \approx n(x|\mathrm{T})$

2. CHECK IF IT IS DIFFERENT FROM THE REFERENCE ONE

$$t(\mathcal{D}) = 2\,\log\left[\frac{e^{-N(\widehat{\mathbf{w}})}}{e^{-N(\mathrm{R})}}\prod_{x\in\mathcal{D}}\frac{n(x|\widehat{\mathbf{w}})}{n(x|\mathrm{R})}\right] \qquad p_{\mathrm{obs}} = \int_{t_{\mathrm{obs}}}^{\infty} dt\, P(t|\mathrm{R})$$

STANDARD LIKELIHOOD RATIO
NEYMAN-PERSON TEST STATISTIC

REFERENCE
DISTRIBUTED
TOYS

THE LOSS FUNCTION

$$n(x|\mathbf{w}) = n(x|\mathrm{R})\, e^{f(x;\mathbf{w})}$$ → NEURAL NETWORK

$$t(\mathcal{D}) = 2\,\log\left[\frac{e^{-N(\widehat{\mathbf{w}})}}{e^{-N(\mathrm{R})}}\prod_{x\in\mathcal{D}}\frac{n(x|\widehat{\mathbf{w}})}{n(x|\mathrm{R})}\right]$$

$$= -2\,\underset{\{\mathbf{w}\}}{\mathrm{Min}}\left[\frac{N(\mathrm{R})}{\mathcal{N}_{\mathcal{R}}}\sum_{x\in\mathcal{R}}\left(e^{f(x;\mathbf{w})}-1\right)-\sum_{x\in\mathcal{D}}f(x;\mathbf{w})\right]$$

$$\equiv -2\,\underset{\{\mathbf{w}\}}{\mathrm{Min}}\,L[f(\,\cdot\,,\mathbf{w})]$$ → LOSS FUNCTION

THE LOSS FUNCTION

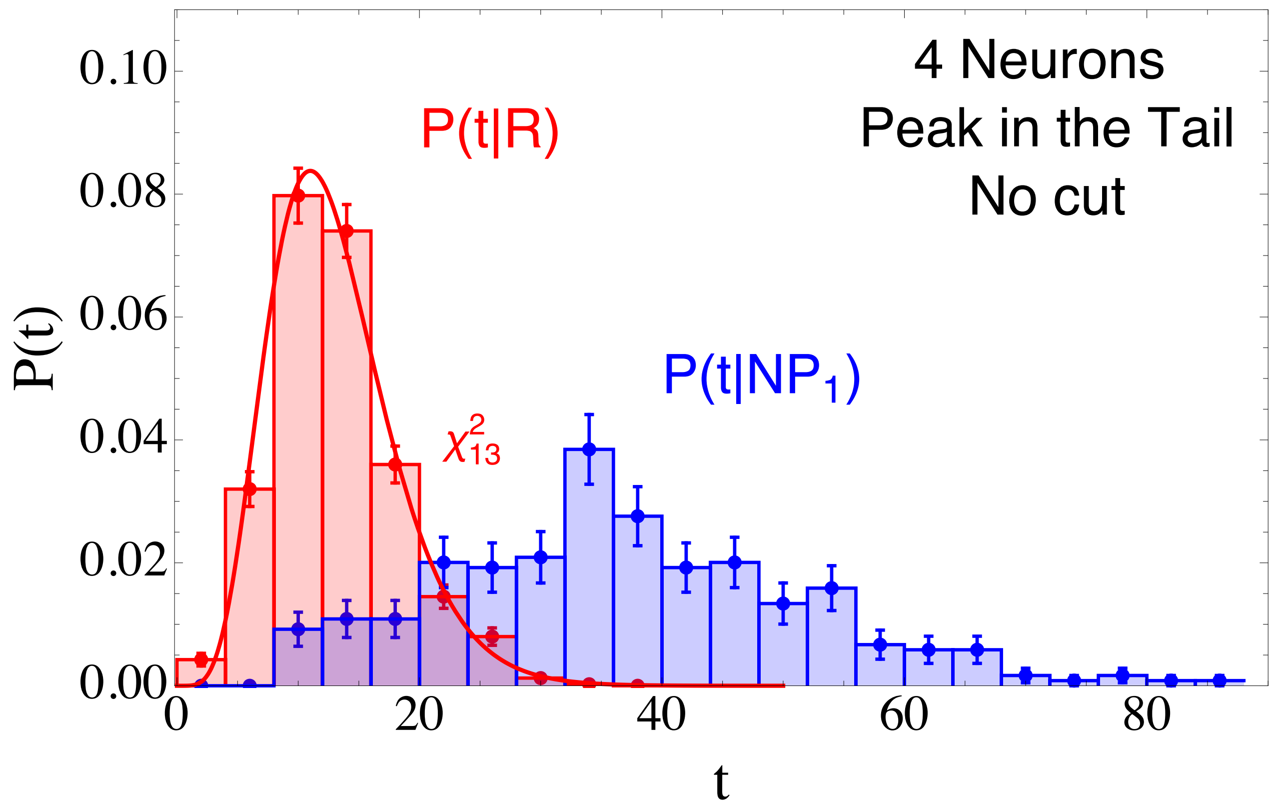
$$n(x|\mathbf{w}) = n(x|\mathrm{R})\, e^{f(x;\mathbf{w})}$$ → NEURAL NETWORK

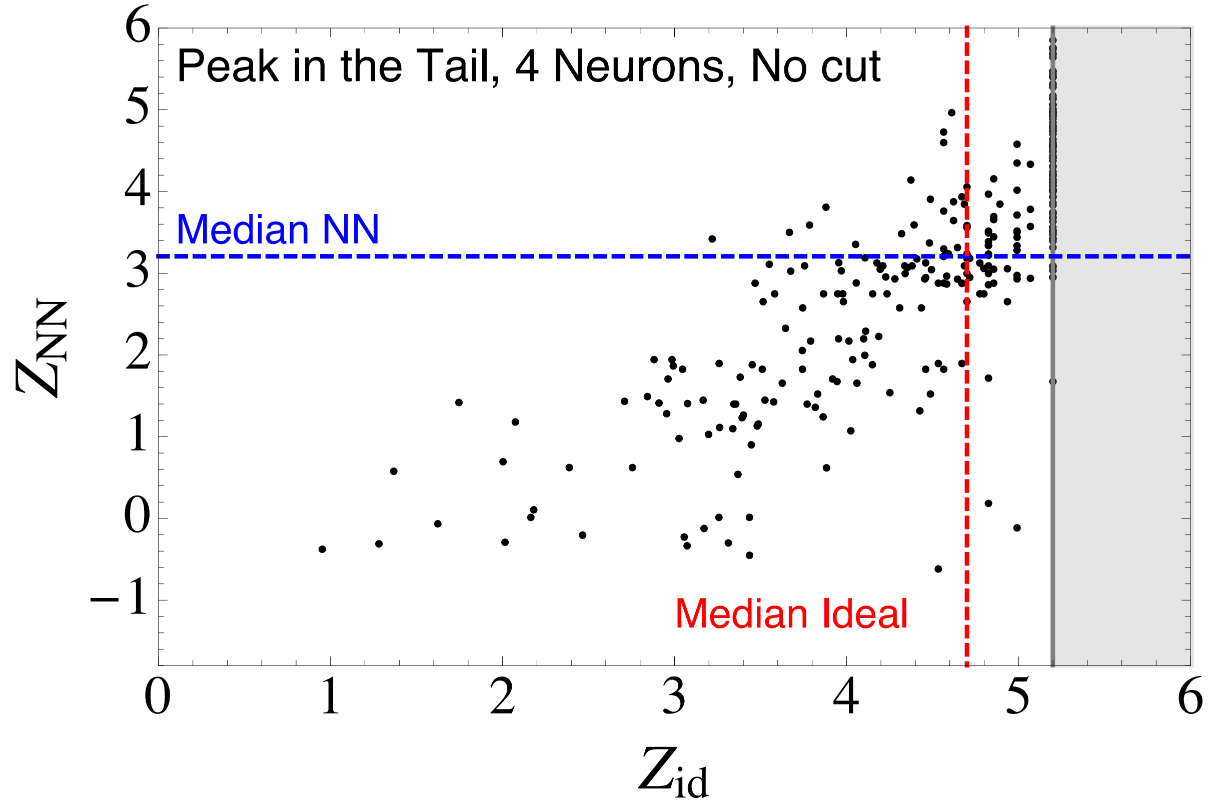
$$t(\mathcal{D}) = 2\log\left[\frac{e^{-N(\widehat{\mathbf{w}})}}{e^{-N(\mathrm{R})}}\prod_{x\in\mathcal{D}}\frac{n(x|\widehat{\mathbf{w}})}{n(x|\mathrm{R})}\right]$$

$$= -2\,\underset{\{\mathbf{w}\}}{\mathrm{Min}}\left[\frac{N(\mathrm{R})}{\mathcal{N}_{\mathcal{R}}}\sum_{x\in\mathcal{R}}\left(e^{f(x;\mathbf{w})}-1\right)-\sum_{x\in\mathcal{D}}f(x;\mathbf{w})\right]$$

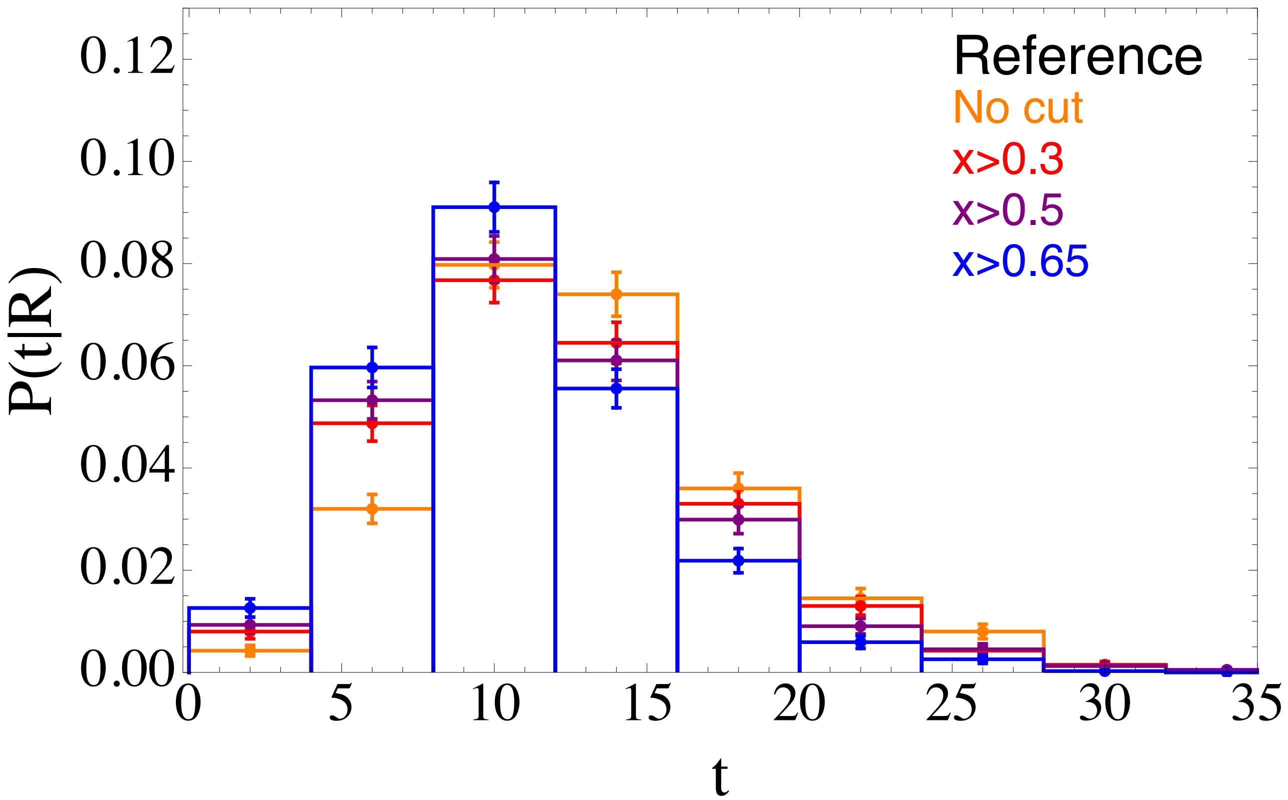
THE NETWORK IS DOING A MAXIMUM LIKELIHOOD FIT TO THE DATA AND COMPUTING THE “OPTIMAL” TEST STATISTIC AT THE SAME TIME

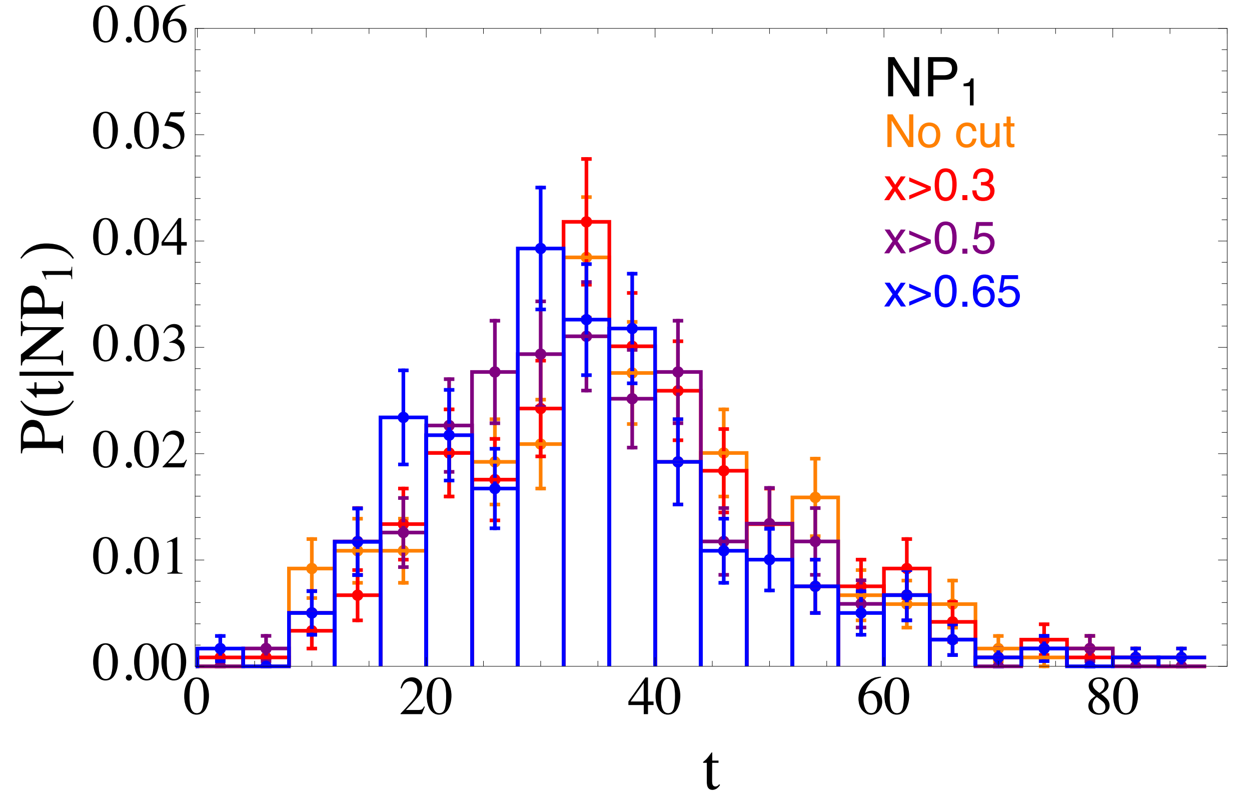
SENSITIVE TO NEW PHYSICS



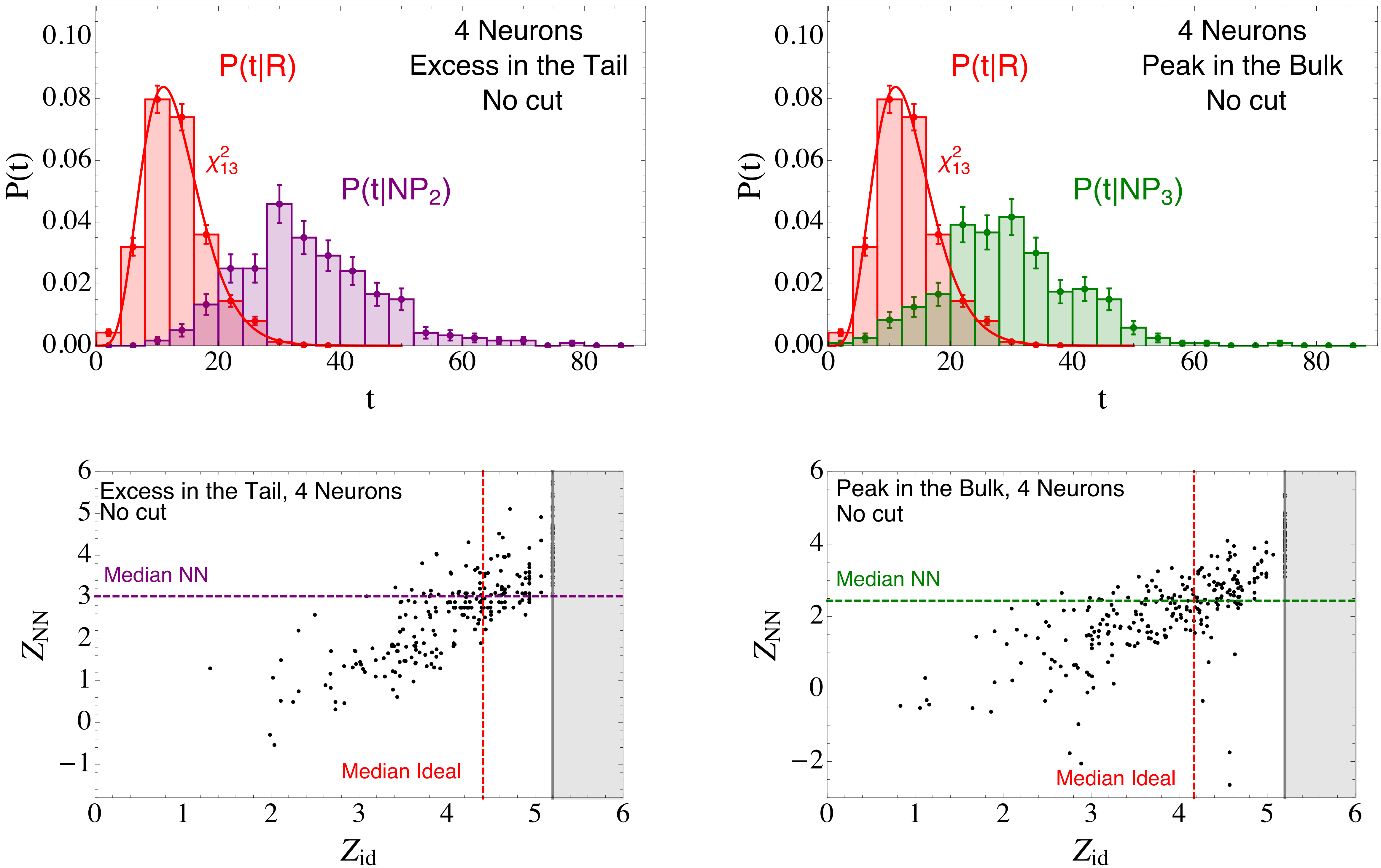


INSENSITIVE TO CUTS

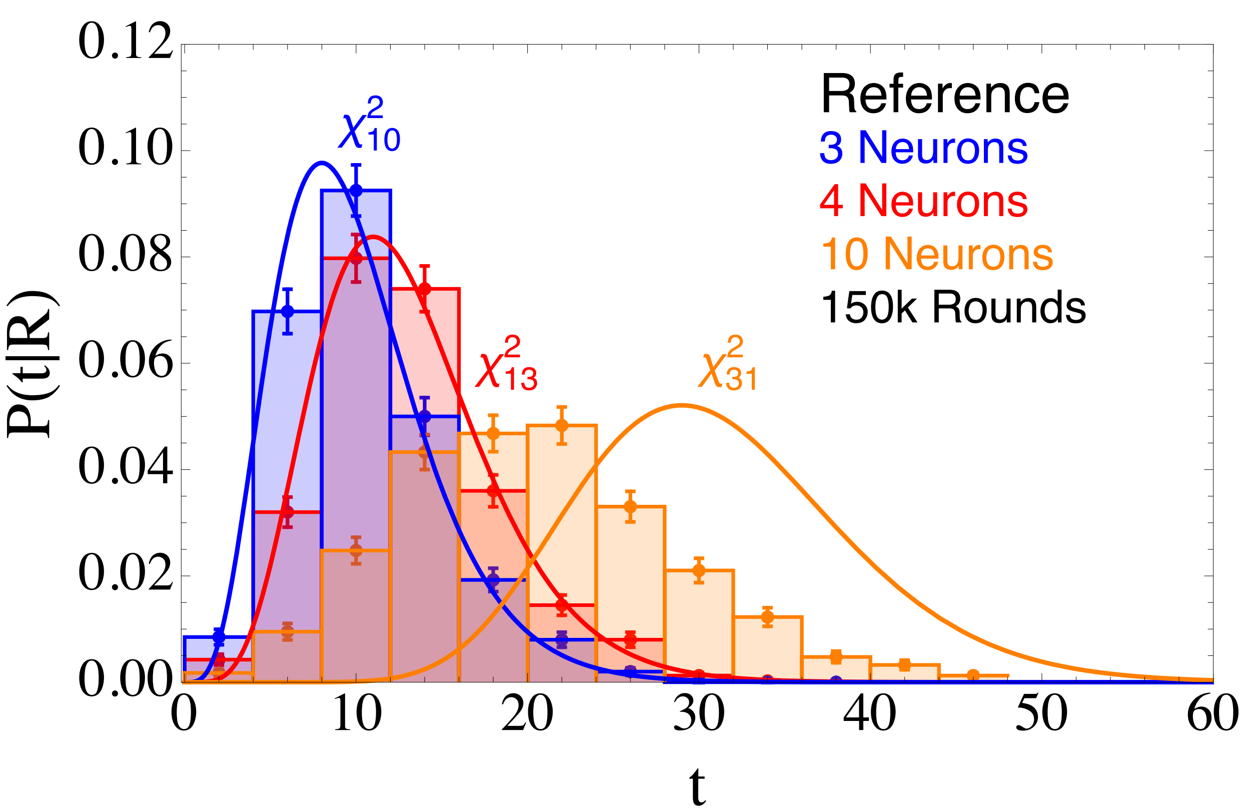


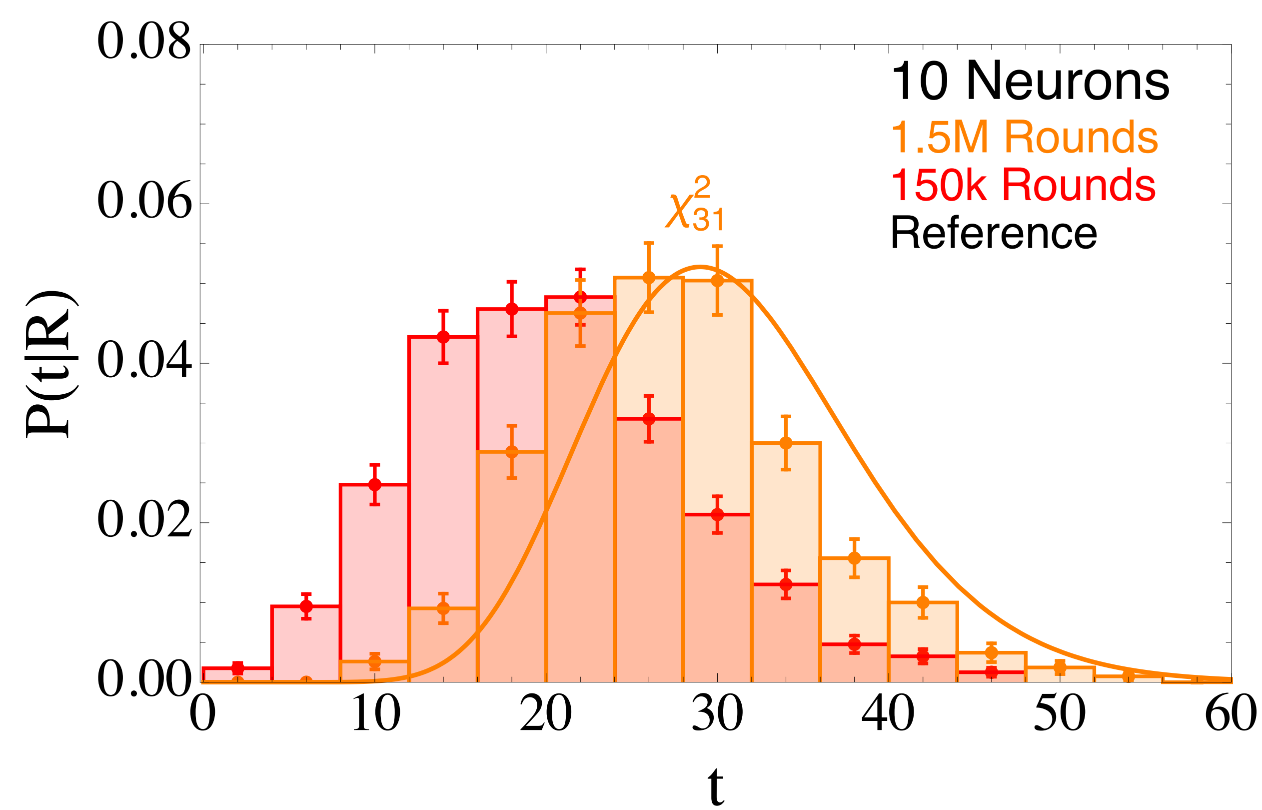


MODEL-INDEPENDENT



NETWORK ARCHITECTURE





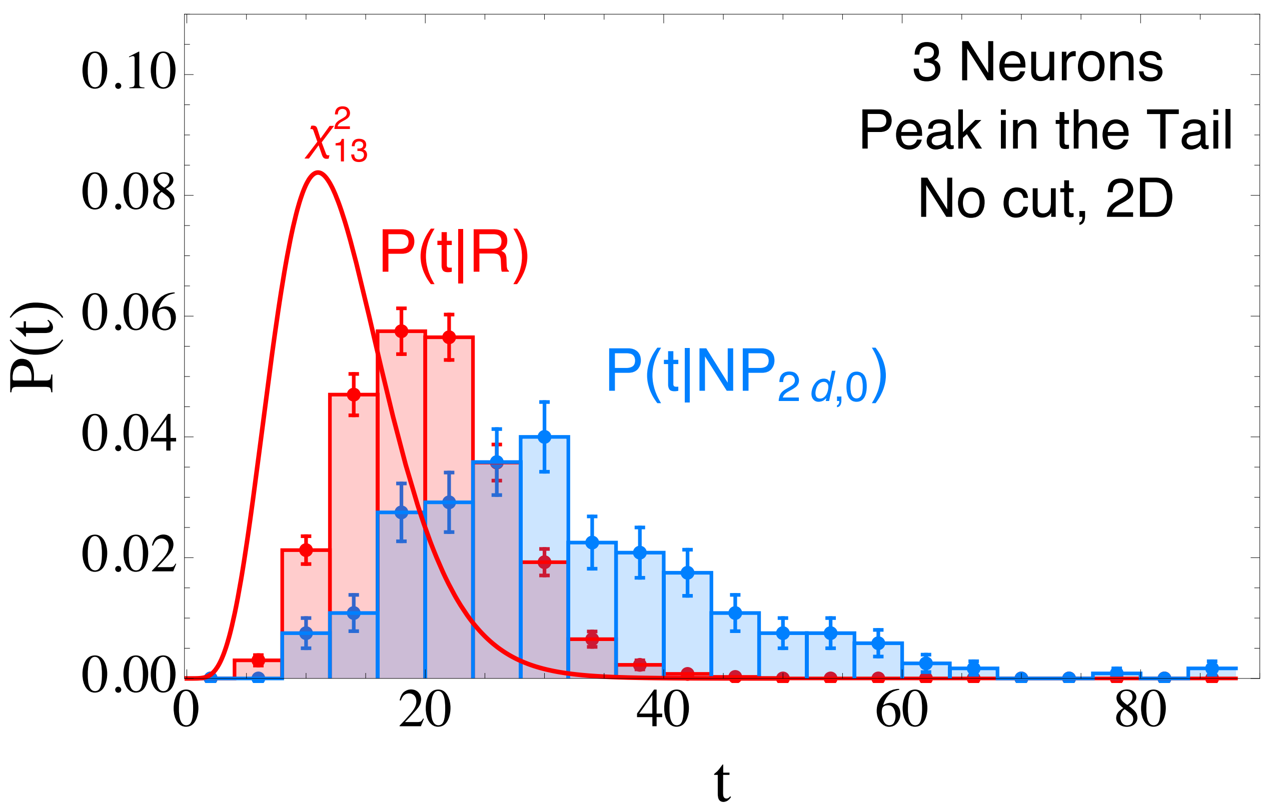
CONCLUSION AND OUTLOOK

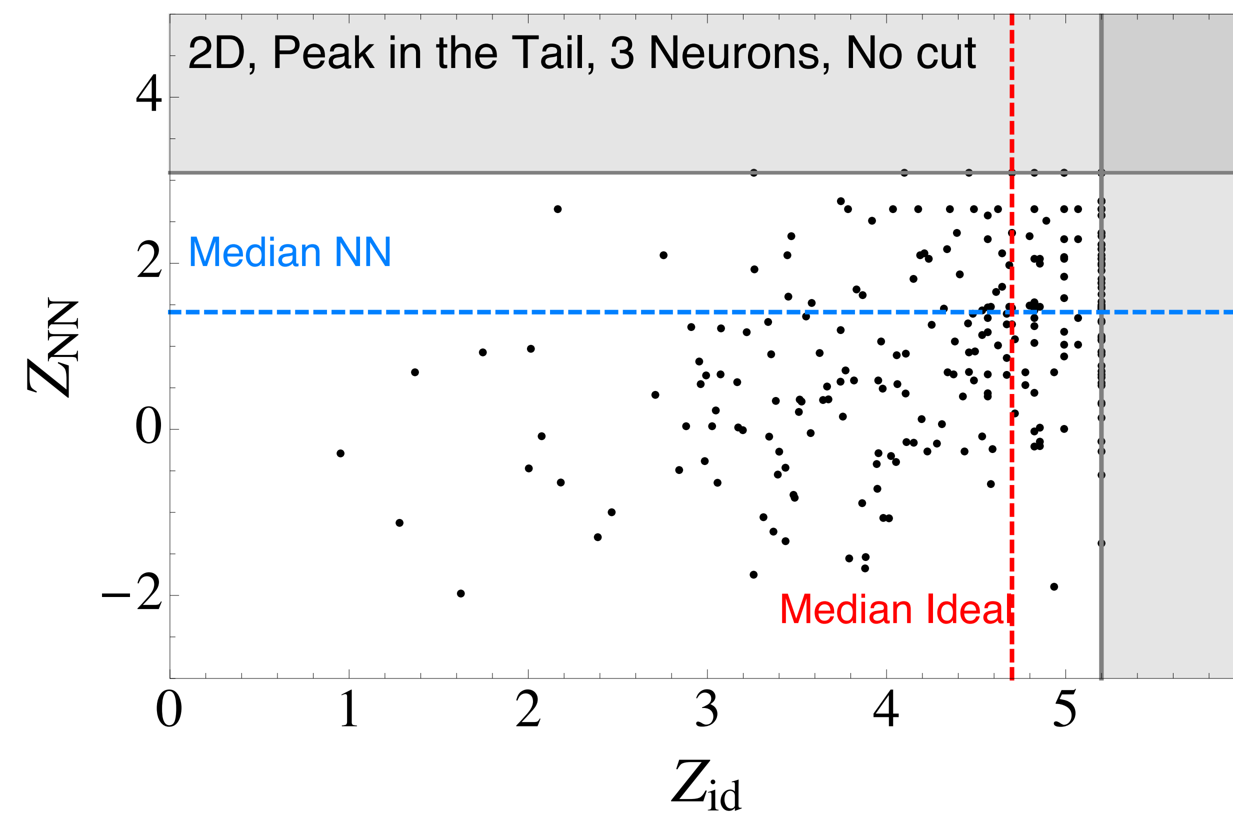
- TODAY IN FUNDAMENTAL PHYSICS WE HAVE LARGE, MULTIVARIATE, SM-LIKE DATASETS AND STRONG REASONS TO BELIEVE THAT THEY SHOULD NOT BE SM-LIKE
- OUR BEST GUESSES FOR NEW PHYSICS ARE NOT BEING DETECTED AND ANYTHING THAT HELPS US TO SEARCH WITHOUT ANY BIAS CAN BE USEFUL
- NEURAL NETWORKS ARE WIDELY USED TO APPROXIMATE PROBABILITY DISTRIBUTIONS AND ARE IDEAL CANDIDATES FOR THIS TYPE OF PROBLEM
- TODAY I HAVE DESCRIBED AN APPLICATION OF NEURAL NETWORKS, FOUNDED ON SOLID STATISTICAL PRINCIPLES, WHICH GOES IN THIS DIRECTION
  - ITS VIRTUES (SENSITIVITY TO NP, MODEL-INDEPENDENCE, INSENSITIVITY TO CUTS) HAVE BEEN TESTED ON SIMPLE 1D AND 2D EXAMPLES
  - MORE WORK IS NEEDED IN THE 2D AND HIGHER-DIMENSIONAL CASE

BACKUP

TWO DIMENSIONS

NP: x~EXPONENTIAL+PEAK y~UNIFORM
R: x~EXPONENTIAL y~UNIFORM





RECOVERS COMPARABLE SENSITIVITY TO 1D FOR x>0.3 OR DOUBLING THE EVENTS