REPRESENTATION OF DISCRETE MAXIMUM ENTROPY DISTRIBUTIONS AS TENSOR NETWORKS

RESEARCH NOTES IN THE ENEXA AND QROM PROJECTS

October 31, 2025

ABSTRACT

We here summarize results of the main report on maximum entropy distributions. The principle of maximum entropy serves as motivation for Computation Activation Networks. We then restrict to these distributions and study relative entropy minimization problems.

In the report, first discussion on the mean polytope are in Chapter 3, and the general maximum entropy problem in Chapter 4. Chapter 8 contains the discussion on maximum entropy distributions in case of boolean statistics.

1 Contents

We in this paper provide tensor network representations

- Representation of any distribution with a sufficient statistics: Generic activation tensors.
- Representation of distributions with maximum entropy, in case of positive realizability: Elementary activation tensors
- Representation of generic maximum entropy distributions: CP activation tensors.

Now, we want to characterize the CP rank of the activation tensors

- Depends on the face of the mean polytope, which contains the mean parameter
- We have thus a well-defined "CP rank" of faces
- Largest faces and vertices have always CP rank of 1, intermediate faces can have larger CP rank

For boolean statistics we further provide insights for boolean statistics (see Chapter 8.5):

- Example of independent statistics (see Exa. 8.28): Always elementary activation tensors (hypercubes)
- Example of partition statistics (see Exa. 8.30):
- Generic criterion for elementary activation: "Cube-like" polytopes (see Def. 8.29)

2 Motivation

Consider a learning problem where we want to estimate a model based on observed data. The maximum entropy problem principle approaches this problem by designing statistics of the data, which means shall be reproduced in the model, and choosing the model reproducing the means of the statistic with least structure. The entropy of a distribution quantifies the degree of structureless in a distribution and is therefore maximized to solve the learning task.

3 The Maximum Entropy Problem

The mean parameter of a distribution $\mathbb{P}\left[X_{[d]}\right]$ to a statistic $\mathcal{S} : \bigtimes_{k\in[d]}[m_k] \to \bigtimes_{s\in[n]}[p_s]$ is the vector $\mu\left[L\right] \in \mathbb{R}^p$ with the coordinates

$$\mu\left[L=l\right] = \mathbb{E}\left[f_l\right] = \left\langle \mathbb{P}\left[X_{[d]}\right], f_l\left[X_{[d]}\right] \right\rangle \left[\varnothing\right] .$$

We express the computation of the mean parameter in the contraction of the selection encoding $\sigma^{\mathcal{S}}\left[X_{[d]}, L\right]$ of $\mathcal{S}$

$$\mu\left[L\right] = \left\langle \mathbb{P}\left[X_{[d]}\right], \sigma^{\mathcal{S}}\left[X_{[d]}, L\right] \right\rangle \left[L\right] .$$

The maximum entropy problem given a mean parameter $\mu^*[L]$ is

$$\max_{\mathbb{P}\left[X_{[d]}\right]\in\Lambda^{\delta,\mathrm{MAX},\nu}} \mathbb{H}\left[\mathbb{P}\left[X_{[d]}\right]\right] \quad \text{subject to} \quad \left\langle \mathbb{P}\left[X_{[d]}\right], \sigma^{\mathcal{S}}\left[X_{[d]}, L\right] \right\rangle \left[L\right] = \mu^*[L]$$

A quick argument shows, that maximum entropy distributions always have $\mathcal{S}$ as a sufficient statistics.

Theorem 1. *Any maximum entropy distribution with respect to a moment constraint on $\mathcal{S}$ and a base measure ν has the sufficient statistic $\mathcal{S}$.*

Proof. Let $\mathbb{P}\left[X_{[d]}\right]$ be a feasible distribution for the maximum entropy problem, which does not have a sufficient statistic $\mathcal{S}$. Then we find $x_{[d]}, \tilde{x}_{[d]} \in \bigtimes_{k\in[d]}[m_k]$ with $x_{[d]} \neq \tilde{x}_{[d]}$, $\mathcal{S}\left(x_{[d]}\right) = \mathcal{S}\left(\tilde{x}_{[d]}\right)$, $\nu\left[X_{[d]} = x_{[d]}\right] \neq 0$, $\nu\left[X_{[d]} = \tilde{x}_{[d]}\right]$ and $\mathbb{P}\left[X_{[d]} = x_{[d]}\right] \neq \mathbb{P}\left[X_{[d]} = \tilde{x}_{[d]}\right]$. We then define a distribution $\tilde{\mathbb{P}}[X_{[d]}]$ coinciding with $\mathbb{P}\left[X_{[d]}\right]$ except for the coordinates $x_{[d]}, \tilde{x}_{[d]}$, where we set

$$\tilde{\mathbb{P}}[X_{[d]} = x_{[d]}] = \tilde{\mathbb{P}}[X_{[d]} = \tilde{x}_{[d]}] = \frac{\mathbb{P}\left[X_{[d]} = x_{[d]}\right] + \tilde{\mathbb{P}}[X_{[d]} = \tilde{x}_{[d]}]}{2}$$

We notice that $\tilde{\mathbb{P}}[X_{[d]}]$ is also a feasible distribution with an larger entropy than $\mathbb{P}\left[X_{[d]}\right]$. Therefore, a distribution which does not have the sufficient statistic $\mathcal{S}$ cannot be a maximum entropy distribution. □

This shows that any maximum entropy distribution is in $\Lambda^{\mathcal{S},\mathrm{MAX}}$, where MAX is the maximal hypergraph $\mathrm{MAX} = ([p], \{[p]\})$. We search for sparse representations of the corresponding activation tensors and investigate in which cases the maximum entropy distribution is also in $\Lambda^{\mathcal{S},\mathcal{G}}$ for sparser hypergraphs $\mathcal{G}$.

4 Preparation

To prepare for the presentation of our main results we introduce

- Computation-Activation Networks: A tensor network architecture, which will be used to represent maximum entropy distributions
- Mean polytopes: Polytopes, which contain all realizable mean parameter vectors.

We will then show, that dependent on the position of the mean parameter in the mean polytope, we can characterize the corresponding maximum entropy distribution by a Computation-Activation Network.

4.1 Computation-Activation Networks

Given a statistic $\mathcal{S} : \bigtimes_{k\in[d]}[m_k] \to \bigtimes_{s\in[n]}[p_s]$ we build its basis encoding tensor

$$\beta^{\mathcal{S}}\left[Y_{[p]}, X_{[d]}\right] = \sum_{x_{[d]}\in\bigtimes_{k\in[d]}[m_k]} \epsilon_{\mathcal{S}(x_{[d]})}\left[Y_{[p]}\right] \otimes \epsilon_{x_{[d]}}\left[X_{[d]}\right] .$$

A computation network is any representation of $\beta^{\mathcal{S}}\left[Y_{[p]}, X_{[d]}\right]$ as a tensor network. These can be constructed in the case statistics being a composition of connective functions.

An activation tensor is $\tau\left[Y_{[p]}\right]$ and the Computation Activation Network of $\mathcal{S}$ and τ the tensor

$$\mathbb{P}\left[X_{[d]}\right] = \left\langle \beta^{\mathcal{S}}\left[Y_{[p]}, X_{[d]}\right], \tau\left[Y_{[p]}\right] \right\rangle \left[X_{[d]}|\varnothing\right] .$$

We are interested in decomposition formats of $\tau\left[Y_{[p]}\right]$, where we use sets of tensor networks $\mathcal{T}^{\mathcal{G}}$ on a hypergraph $\mathcal{G}$. The family of by $\mathcal{S}$ and a $\mathcal{G}$ computable distributions are

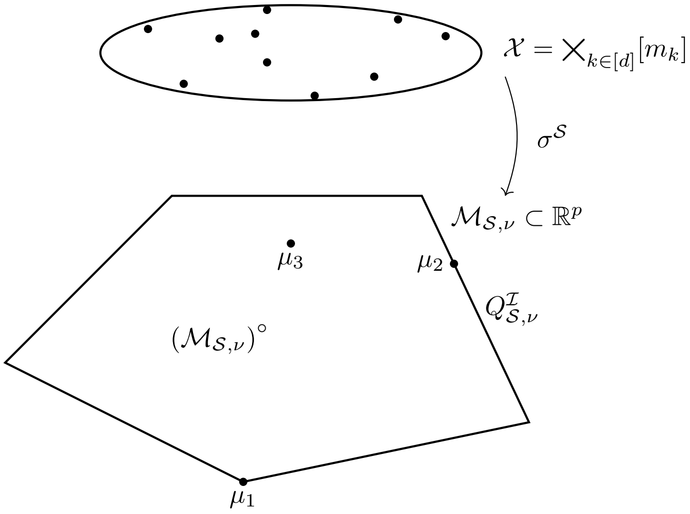
$$\Lambda^{\mathcal{S},\mathcal{G}} = \left\{ \left\langle \beta^{\mathcal{S}}\left[Y_{[p]}, X_{[d]}\right], \tau\left[Y_{\mathcal{V}}\right] \right\rangle \left[X_{[d]}|\varnothing\right] \,:\, \tau\left[Y_{\mathcal{V}}\right] \in \mathcal{T}^{\mathcal{G}} \right\} .$$

4.2 The mean polytope

The mean polytope is the set of mean parameters to any distribution. We define it

$$\mathcal{M}_{\mathcal{S},\nu} = \left\{ \left\langle \mathbb{P}, \sigma^{\mathcal{S}}, \nu \right\rangle [L] \,:\, \mathbb{P}\left[X_{[d]}\right] \in \Lambda^{\delta,\mathrm{MAX},\nu} \right\} ,$$

where we denote by $\Lambda^{\delta,\mathrm{MAX},\nu}$ the set of all probability distributions representable with respect to ν.


$\mathcal{X} = \bigtimes_{k\in[d]}[m_k]$
$\sigma^{\mathcal{S}}$
$\mathcal{M}_{\mathcal{S},\nu} \subset \mathbb{R}^p$
μ_3
μ_2
$Q^{\mathcal{I}}_{\mathcal{S},\nu}$
$(\mathcal{M}_{\mathcal{S},\nu})^\circ$
μ_1

The mean polytope is the convex hull

$$\mathcal{M}_{\mathcal{S},\nu} = \mathrm{conv}\left(\sigma^{\mathcal{S}}\left[X_{[d]} = x_{[d]}, L\right] \,:\, x_{[d]} \in \bigtimes_{k\in[d]} [m_k],\, \nu\left[X_{[d]} = x_{[d]}\right] = 1 \right) .$$

It is thus a convex polytope, inherited by the convex polytope of distributions (the standard simplex). We can characterize the maximum entropy distribution based on the position of the mean parameter in the mean polytope. To be more precise, any polytope decomposes into effective interiors of its faces and we characterize the maximum entropy distribution depending on the face to the mean parameter.

To be included:

- Faces of the mean polytope are itself mean polytopes with respect to refined base measures.
- Existence of vectors in the image of the statistic encoding corresponds with satisfiability of the corresponding formula.

5 Tensor network representation of maximum entropy distributions

Given the mean polytope discussion we now characterize the tensor network representation of maximum entropy distributions.

5.1 Maximum entropy on the interior

A classical result states, that the maximum entropy distribution is in the exponential family $\Gamma^{\mathcal{S},\nu}$.

Theorem 2. *If and only if μ^* is in the effective interior of $\mathcal{M}_{\mathcal{S},\nu}$, then the unique solution of the maximum entropy problem is the distribution*

$$\mathbb{P}^{\mathcal{S},\mu^*,\nu}[X_{[d]}] \in \Gamma^{\mathcal{S},\nu}$$

with $\left\langle \mathbb{P}^{\mathcal{S},\mu^*,\nu}[X_{[d]}], \sigma^{\mathcal{S}}\left[X_{[d]}, L\right] \right\rangle [L] = \mu^*[L]$.

Proof. By The 3.3 in [Wainwright and Jordan, 2008], since by assumption

$$\mu\left[L\right] \in \left(\mathcal{M}_{\mathcal{S},\nu}\right)^{\circ} ,$$

there is a canonical parameter θ with

$$\left\langle \mathbb{P}^{\mathcal{S},\theta,\nu}[X_{[d]}], \sigma^{\mathcal{S}}\left[X_{[d]}, L\right] \right\rangle \left[L\right] = \mu\left[L\right] .$$

For any other feasible distribution $\tilde{\mathbb{P}}[X_{[d]}]$ we also have $\left\langle \tilde{\mathbb{P}}[X_{[d]}], \sigma^{\mathcal{S}}\left[X_{[d]}, L\right] \right\rangle \left[L\right] = \mu\left[L\right]$ and thus

$$\begin{aligned}
\mathbb{H}\left[\tilde{\mathbb{P}}, \mathbb{P}^{(\mathcal{S},\theta,\nu)}\right] &= -\left\langle \tilde{\mathbb{P}}, \ln\left[\mathbb{P}^{(\mathcal{S},\theta,\nu)}[X_{[d]}]\right] \right\rangle \left[\varnothing\right] \\
&= -\left\langle \tilde{\mathbb{P}}, \left\langle \sigma^{\mathcal{S}}\left[X_{[d]}, L\right], \theta\left[L\right] \right\rangle \left[X_{[d]}\right] \right\rangle \left[\varnothing\right] + A^{(\mathcal{S},\nu)}(\theta) \\
&= -\left\langle \theta, \mu \right\rangle \left[\varnothing\right] + A^{(\mathcal{S},\nu)}(\theta) \\
&= \mathbb{H}\left[\mathbb{P}^{(\mathcal{S},\theta,\nu)}\right] .
\end{aligned}$$

With the Gibbs inequality we have if $\tilde{\mathbb{P}} \neq \mathbb{P}^{(\mathcal{S},\theta,\nu)}$

$$\mathbb{H}\left[\mathbb{P}^{(\mathcal{S},\hat{\theta},\nu)}\right] - \mathbb{H}\left[\tilde{\mathbb{P}}\right] = \mathbb{H}\left[\tilde{\mathbb{P}}, \mathbb{P}^{(\mathcal{S},\hat{\theta},\nu)}\right] - \mathbb{H}\left[\tilde{\mathbb{P}}\right] > 0 .$$

Therefore, if $\tilde{\mathbb{P}}$ does not coincide with$\mathbb{P}^{(\mathcal{S},\hat{\theta},\nu)}$, it is not a maximum entropy distribution. □

Exponential families are in $\Lambda^{\mathcal{S},\mathrm{EL}}$, if and only if $\left\langle \nu \right\rangle \left[X_{[d]}|\varnothing\right] \in \Lambda^{\mathcal{S},\mathrm{EL}}$. If $\left\langle \nu \right\rangle \left[X_{[d]}|\varnothing\right] \in \Lambda^{\mathcal{S},\mathrm{EL}}$ and $\mu\left[L\right] \in \left(\mathcal{M}_{\mathcal{S},\nu}\right)^{\circ}$ we therefore have a sparse representation of the maximum entropy distribution with elementary activation tensors.

5.2 Mean parameter on faces

We always find a unique face of the polytope with the mean parameter being in the interior. Any distribution reproducing the mean parameter is realizable with respect to the face measure of that face. We conclude that the maximum entropy distribution of $\mu^*[L]$ with respect to $\mathcal{S}, \nu$ is also the maximum entropy distribution

Theorem 3. *Given $\mathcal{S}$ and $\mu\left[L\right] \in \mathcal{M}_{\mathcal{S},\nu}$, let $\mathcal{I}$ be the smallest face of $\mathcal{M}_{\mathcal{S},\nu}$ such that*

$$\mu\left[L\right] \in Q^{\mathcal{I}}_{\mathcal{S},\nu} .$$

Then the corresponding maximum entropy distribution is in $\Lambda^{\mathcal{S},\mathcal{G},\nu}$ if and only if the face measure (see Def. **??***)*

$$\kappa^{\mathcal{I}}\left[Y_{[p]}\right] = \sum_{\mu \in Q^{\mathcal{I}}_{\mathcal{S},\nu} \cap \mathrm{im}(\sigma^{\mathcal{S}})} \epsilon_{\mu}\left[Y_{[p]}\right]$$

is in $\mathcal{T}^{\mathcal{G}}$.

Proof. By Thm. **??** the maximum entropy distribution is an element of the exponential family with by the face measure refined base measure $\tilde{\nu}$. Let $\theta\left[L\right]$ be a canonical parameter such that

$$\left\langle \sigma^{\mathcal{S}}\left[X_{[d]}, L\right], \mathbb{P}^{\mathcal{S},\theta,\tilde{\nu}}\left[X_{[d]}\right] \right\rangle \left[L\right] = \mu\left[L\right] ,$$

that is $\mathbb{P}^{\mathcal{S},\theta,\tilde{\nu}}\left[X_{[d]}\right]$ is the maximum entropy distribution. We apply Thm. **??** to represent the face measure by

$$\nu^{\mathcal{S},\mathcal{I}}\left[X_{[d]}\right] = \left\langle \beta^{\mathcal{S}}\left[Y_{[p]}, X_{[d]}\right], \kappa^{\mathcal{I}}\left[Y_{[p]}\right] \right\rangle \left[X_{[d]}\right]$$

Then for the tensor

$$\tau\left[Y_{[p]}\right] = \left\langle \left\{\alpha^{l,\theta}\left[Y_l\right] \,:\, l \in [p]\right\} \cup \left\{\kappa^{\mathcal{I}}\left[Y_{[p]}\right]\right\} \right\rangle \left[Y_{[p]}\right]$$

we have

$$\mathbb{P}^{\mathcal{S},\theta,\tilde{\nu}}\left[X_{[d]}\right] = \left\langle \beta^{\mathcal{S}}\left[Y_{[p]}, X_{[d]}\right], \tau\left[Y_{[p]}\right], \nu\left[X_{[d]}\right] \right\rangle \left[X_{[d]}|\varnothing\right] .$$

Thus, the maximum entropy distribution is in $\Lambda^{\mathcal{S},\mathcal{G},\nu}$, if τ admits a tensor network decomposition with respect to $\mathcal{G}$. Since the hard activation cores are elementary, this is the case when $\kappa^{\mathcal{I}}$ admits a tensor network decomposition with respect to $\mathcal{G}$. □

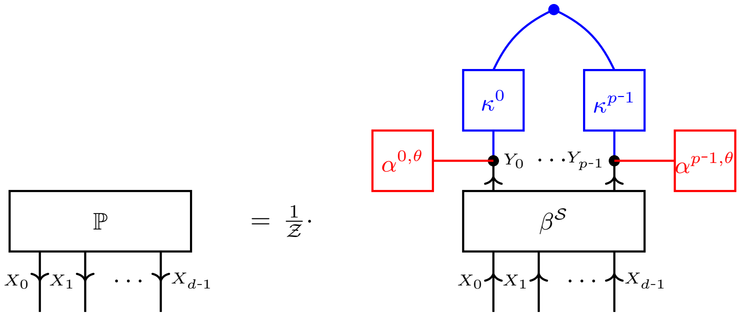


Figure 1: Tensor network decomposition of maximum entropy distributions to the constraint $\mu[L] = \langle \mathbb{P}, \sigma^{\mathcal{S}} \rangle [L]$. Blue: Constraint activation cores κ^l in a CP decomposition, representing the face measure to the minimal face, such that $\mu \in Q^{\mathcal{I}}_{\mathcal{S},\nu}$. Red: Probabilistic activation cores $\alpha^{l,\theta}[Y_l]$ in an elementary decomposition, where each leg core is a scaled exponentials evaluated on the enumerated image $\mathrm{im}(s_l)$.

6 Characterization for boolean statistics

For boolean statistics $\mathcal{F} : \times_{k\in[d]}[m_k] \to \times_{l\in[p]}[2]$ the mean polytope is a subset of the cube $[0,1]^p$. In this case, any boolean vector in $\mathcal{M}_{\mathcal{F},\nu}$ is a vertex. It follows, that any distribution reproducing a mean parameter $\mu[L]$ on the effective interior of $\mathcal{M}_{\mathcal{F},\nu}$ is positive with respect to ν.

We apply the exponential distribution characterization of the maximum entropy distribution and get that the maximum entropy distribution is in $\Lambda^{\mathcal{S},\mathrm{EL}}$, if and only if the face measure is in $\Lambda^{\mathcal{S},\mathrm{EL}}$. This is exactly the case, when the face is an intersection of the mean polytope with a face of the cupe $[0,1]^p$.

The mean parameters, which can be realized by a distribution in $\Lambda^{\mathcal{F},\mathrm{EL}}$ are those, which are on the effective interior of the intersection of the mean polytope with a face of the cube.

6.1 Example

Example 1 (Maximum entropy distribution with non-elementary activation cores)**.** *Consider two atomic variables X_0 and X_1 and a statistic $\mathcal{F}$ consisting in the formulas*

$$f_0 = (X_0 \land X_1) \quad , \quad f_1 = (X_0 \Rightarrow X_1)$$

with the coordinatewise expressions

$$f_0 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad , \quad f_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} .$$

We can think of X_0 as a feature on an invoice, and X_1 as a feature on the accounting proposal.

From this we have

$$\beta^{(f_0,f_1)}[Y_0 = 0, Y_1 = 0, X_0, X_1] = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \quad , \quad \beta^{(f_0,f_1)}[Y_0 = 0, Y_1 = 1, X_0, X_1] = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \quad ,$$
$$\beta^{(f_0,f_1)}[Y_0 = 1, Y_1 = 0, X_0, X_1] = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \beta^{(f_0,f_1)}[Y_0 = 1, Y_1 = 1, X_0, X_1] = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} .$$

Since the only vanishing slice of $\beta^{\mathcal{F}}$ with respect to the head variables is that to $y_{0,1} = (1,0)$, the vertices of the mean polytope are the vectors to the other head indices. The mean polytope is the convex hull of these vertices

$$\mathcal{M}_{(f_0,f_1)} = \mathrm{conv}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) .$$

This polytope has a non cube-like face (sketched blue in Figure 2), which is the convex hull of the vertices $[0\,0]^T$, $[1\,1]^T$. This face is parametrized by the (CP*-rank 2) hard activation core*

$$\kappa^{(0,0),(1,1)}[Y_0, Y_1] = \epsilon_{(0,0)}[Y_0, Y_1] + \epsilon_{(1,1)}[Y_0, Y_1] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

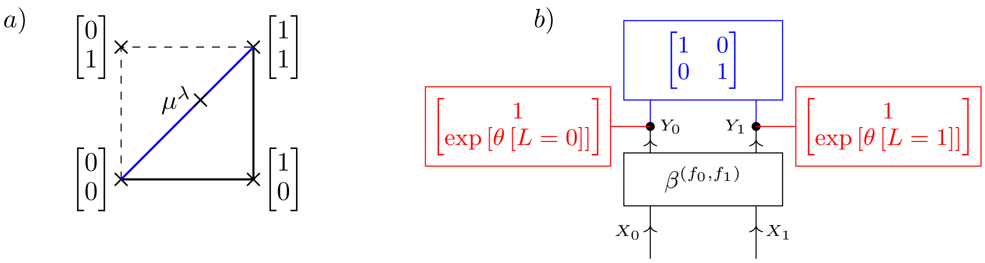

a)
$\begin{bmatrix}0\\1\end{bmatrix}$
$\begin{bmatrix}1\\1\end{bmatrix}$
μ^λ
$\begin{bmatrix}0\\0\end{bmatrix}$
$\begin{bmatrix}1\\0\end{bmatrix}$
b)
$\begin{bmatrix}1&0\\0&1\end{bmatrix}$
$\begin{bmatrix}1\\\exp\left[\theta\left[L=0\right]\right]\end{bmatrix}$
Y_0
Y_1
$\begin{bmatrix}1\\\exp\left[\theta\left[L=1\right]\right]\end{bmatrix}$
$\beta^{(f_0,f_1)}$
X_0
X_1

Figure 2: a) Mean polytope of the statistic $\mathcal{F} = (X_0 \wedge X_1, X_0 \Rightarrow X_1)$ (thick), as a subset of the cube $[0,1]^2$ (dashed). The blue line is the face of the polytope, which is not cube like, that is not an intersection of the polytope with the faces of the polytope. We further define for $\lambda \in (0,1)$ a mean parameter $\mu_\lambda\left[L\right] = \left[\lambda\, \lambda\right]^T$ which is on the interior of the blue face. b) Corresponding Computation-Activation Network being the maximum entropy distribution reproducing $\mu_\lambda\left[L\right]$, when λ is the sigmoid of $\theta\left[L=0\right] + \theta\left[L=1\right]$.

and has the face measure

$$\left\langle \kappa^{(0,0),(1,1)}\left[Y_0,Y_1\right], \beta^{\mathcal{F}}\left[Y_0,Y_1,X_0,X_1\right]\right\rangle\left[X_0,X_1\right] = \begin{bmatrix}0&0\\1&1\end{bmatrix}.$$

Any mean parameter μ on the interior of that face can be parametrized by a scalar $\lambda \in (0,1)$

$$\mu_\lambda\left[L\right] = \begin{bmatrix}\lambda & \lambda\end{bmatrix}^T .$$

With the canonical parameters $\theta\left[L\right] \in \mathbb{R}^2$ of the maximum entropy distributions on this face by

$$\mathbb{P}\left[X_0,X_1\right] = \frac{1}{1+\exp\left[\theta\left[L=0\right]+\theta\left[L=1\right]\right]}\begin{bmatrix}0 & 0\\1 & \exp\left[\theta\left[L=0\right]+\theta\left[L=1\right]\right]\end{bmatrix}$$

we get the correspondence by the sigmoid

$$\lambda = \frac{1}{1+\exp\left[-(\theta\left[L=0\right]+\theta\left[L=1\right])\right]} .$$

Note, that the hard activation core $\kappa^{(0,0),(1,1)}\left[Y_0,Y_1\right]$ to the blue face is the only non-elementary activation core. While the vertices have always elementary cores, the further non-vertex faces have elementary activation cores

$$\kappa^{(0,0),(1,0),(1,1)}\left[Y_0,Y_1\right] = \begin{bmatrix}1&1\\1&1\end{bmatrix} = \mathbb{I}\left[Y_0\right]\otimes\mathbb{I}\left[Y_1\right] \quad , \quad \kappa^{(0,0),(1,0)}\left[Y_0,Y_1\right] = \begin{bmatrix}1&0\\1&0\end{bmatrix} = \mathbb{I}\left[Y_0\right]\otimes\epsilon_0\left[Y_1\right] \quad ,$$
$$\kappa^{(1,0),(1,1)}\left[Y_0,Y_1\right] = \begin{bmatrix}1&1\\0&0\end{bmatrix} = \epsilon_0\left[Y_0\right]\otimes\mathbb{I}\left[Y_1\right] .$$

The maximum entropy distributions to mean parameters on the interior of all other faces than the blue face are represented by Computation-Activation Networks with only elementary activation cores.