Characterization of Computation-Activation Networks by sufficient statistics

Research Notes in the ENEXA and QROM Projects

November 14, 2025

1 Sufficient Statistic for Parametrized Families

Sufficient statistics are treated in mathematical statistics and in information theory. We here choose a definition of information theory and apply a factorization theorem of mathematical statistics to relate with Computation-Activation Networks. The distribution of a canonical parameter is now drawn from a (possibly continuous) random variable Θ, which takes values $\theta \in \Gamma$ with probability

$$\tilde{\mathbb{P}}[\Theta = \theta] .$$

Definition 1 (Sufficient statistics for Parameters)**.** *Let $\{\mathbb{P}^\theta \left[X_{[d]}\right] \,:\, \theta \in \Gamma\}$ be a family of probability distributions and*

$$\mathcal{S} \,:\, \bigtimes_{k\in[d]} [m_k] \rightarrow \bigtimes_{l\in[p]} [p_l]$$

be a function. We say that $\mathcal{S}$ is sufficient for Θ, if for any distribution $\tilde{\mathbb{P}}[\Theta]$ of Θ, when drawing $X_{[d]}$ from $\mathbb{P}^\theta \left[X_{[d]}\right]$ with probability $\tilde{\mathbb{P}}[\Theta = \theta]$, we have that

$$\left(\Theta \perp X_{[d]}\right) \big| \, \mathcal{S}\left(X_{[d]}\right) .$$

Theorem 1 (Characterization of Computation-Activation Networks)**.** *Let $\{\mathbb{P}^\theta \left[X_{[d]}\right] \,:\, \theta \in \Gamma\}$ be a family of probability distributions with a sufficient statistic $\mathcal{S}$. Then there is a non-negative (possibly non-Boolean) base measure $\nu\left[X_{[d]}\right]$ and a map*

$$h : \Gamma \rightarrow \bigotimes_{l\in[p]} \mathbb{R}^{p_l}$$

such that for all $\theta \in \Gamma$

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

We further have that for a set $\{\mathbb{P}^\theta \left[X_{[d]}\right] \,:\, \theta \in \Gamma\}$ $\mathcal{S}$ is a sufficient statistic, if and only if there is a non-negative (possibly non-Boolean) base measure $\nu\left[X_{[d]}\right]$ with

$$\{\mathbb{P}^\theta \left[X_{[d]}\right] \,:\, \theta \in \Gamma\} \subset \Lambda^{\mathcal{S},\mathrm{MAX},\nu} .$$

Proof. By the factorization theorem of mathematical statistics (see [Hogg - The. 2.7.1]) we have that $\mathcal{S}$ is a sufficient statistic if and only if there are real-valued functions k_1 on $\left(\bigtimes_{l\in[p]}[p_l]\right) \times \Gamma$ and k_2 on $\bigtimes_{k\in[d]}[m_k]$ such that

$$\mathbb{P}^\theta \left[X_{[d]} = x_{[d]}\right] = k_1(\mathcal{S}^{x_{[d]}}, \Gamma) \cdot k_2(x_{[d]}) . \tag{1}$$

We define a base measure by the coordinate encoding of k_2 by

$$\nu\left[X_{[d]}\right] = \sum_{x_{[d]} \in \times_{k\in[d]}[m_k]} k_2(x_{[d]})\epsilon_{x_{[d]}}\left[X_{[d]}\right]$$

and for each $\theta \in \Gamma$ an activation tensor

$$\xi^{\theta}\left[Y_{[p]}\right] = \sum_{y_{[p]}} k_1(y_{[p]}, \theta)\epsilon_{y_{[p]}}\left[Y_{[p]}\right] .$$

With this we have for any $\theta \in \Gamma$

$$\left\langle h(\Gamma)[Y_{[p]}], \beta^{\mathcal{S}}\left[Y_{[p]}, X_{[d]}\right], \nu\left[X_{[d]}\right]\right\rangle[\varnothing] = 1$$

and thus for any $x_{[d]} \in \times_{k\in[d]}[m_k]$ applying basis calculus

$$\begin{aligned}\left\langle h(\Gamma)[Y_{[p]}], \beta^{\mathcal{S}}\left[Y_{[p]}, X_{[d]}\right], \nu\left[X_{[d]}\right]\right\rangle\left[X_{[d]} = x_{[d]}|\varnothing\right] &= h(\Gamma)[Y_{[p]} = \mathcal{S}^{x_{[d]}}] \cdot \nu\left[X_{[d]} = x_{[d]}\right] \\ &= k_1(\mathcal{S}^{x_{[d]}}, \Gamma) \cdot k_2(x_{[d]}) \\ &= \mathbb{P}^{\theta}\left[X_{[d]} = x_{[d]}\right] .\end{aligned}$$

We therefore find for any $\mathbb{P}^{\theta}\left[X_{[d]}\right]$ a representation as a Computation-Activation Network in $\Lambda^{\mathcal{S},\mathrm{MAX},\nu}$ with the activation tensor $h(\Gamma)[Y_{[p]}]$.

To show the second claim, we are left to show that any set of Computation-Activation Networks in $\Lambda^{\mathcal{S},\mathrm{MAX},\nu}$ has $\mathcal{S}$ as a sufficient statistic. Let us thus consider a parametric family

$$\{\mathbb{P}^{\theta}\left[X_{[d]}\right] \,:\, \theta \in \Gamma\} \subset \Lambda^{\mathcal{S},\mathrm{MAX},\nu} .$$

By this inclusion we find for any $\theta \in \Gamma$ an activation core $\alpha^{\theta}[Y_{[p]}]$. We then construct functions k_1 and k_2 by

$$k_1(y_{[p]}, \Gamma) = \alpha^{\theta}[Y_{[p]} = y_{[p]}] \quad \text{and} \quad k_2(x_{[d]}) = \nu\left[X_{[d]} = x_{[d]}\right]$$

and notice that the equivalent condition (1) to $\mathcal{S}$ being a sufficient statistic is satisfied. □

2 Sufficient Statistic for the Probability

We here consider sufficient statistics for the parameter of a parametrized family, while in the report we considered sufficient statistics for the probability mass as a random variable. In both cases this results from the information theoretic viewpoint, that a function T of X is a sufficient statistic for a variable Z, if

$$(Z \perp X)|\, T(X) .$$

While we choose for Z Y_{θ} above, we now choose for Z the variable $Y_{\mathbb{P}}$. This variable can be computed by contraction with

$$\beta^{\mathbb{P}}\left[Y_{\mathbb{P}}, X_{[d]}\right] .$$

If T is a sufficient statistic for $Y_{\mathbb{P}}$, we call it probability sufficient for $\mathbb{P}$.

Theorem 2 (Theorem 2.19 in the report)**.** *If and only if a statistic $\mathcal{S}$ is probability sufficient for $\mathbb{P}\left[X_{[d]}\right]$, then*

$$\mathbb{P}\left[X_{[d]}\right] \in \Lambda^{\mathcal{S},\mathrm{MAX},\mathbb{I}} .$$

Note that by this theorem w can restrict ourselves to the Computation-Activation Networks with trivial base measure for the characterization of distributions with a probability sufficient statistic.