Why We Can Describe A Complex System's States With Exponential Family

Zhenyu Yang

August 2024

1 The Theoretical Framework

Suppose we have a probabilistic model to describe the system's states ($\mathbf{X}$) under given conditions ($\mathbf{C}$). Here we given the theoretical framework to derive the conditional probability density function $f(\mathbf{X} \mid \mathbf{C})$ by maximizing the conditional entropy.

First, start with the Lagrangian function $\mathcal{L}$:

$$\begin{aligned}\mathcal{L} = \int \Bigg[&- \int_S f(\mathbf{X} \mid \mathbf{C}) \ln f(\mathbf{X} \mid \mathbf{C})\, d\mathbf{X} + \lambda(\mathbf{C}) \left(1 - \int_S f(\mathbf{X} \mid \mathbf{C})\, d\mathbf{X} \right) \\ &+ \sum_{i=1}^{n} \mu_i(\mathbf{C}) \left(\mathbf{g}_i(\mathbf{C}) - \int_S \alpha_i(\mathbf{X}) f(\mathbf{X} \mid \mathbf{C})\, d\mathbf{X} \right) \Bigg] d\pi(\mathbf{C}).\end{aligned} \tag{1}$$

Where α are measurements and g are the expectations of the measurements.

To perform the variation with respect to $f(\mathbf{X} \mid \mathbf{C})$, let $\delta f(\mathbf{X} \mid \mathbf{C})$ be a small variation in $f(\mathbf{X} \mid \mathbf{C})$. We need to compute $\delta\mathcal{L}$:

$$\begin{aligned}\delta\mathcal{L} = \int \Bigg[&- \int_S \delta f(\mathbf{X} \mid \mathbf{C}) \ln f(\mathbf{X} \mid \mathbf{C})\, d\mathbf{X} - \int_S f(\mathbf{X} \mid \mathbf{C}) \frac{\delta f(\mathbf{X} \mid \mathbf{C})}{f(\mathbf{X} \mid \mathbf{C})}\, d\mathbf{X} \\ &+ \lambda(\mathbf{C}) \left(- \int_S \delta f(\mathbf{X} \mid \mathbf{C})\, d\mathbf{X} \right) + \sum_{i=1}^{n} \mu_i(\mathbf{C}) \left(- \int_S \alpha_i(\mathbf{X}) \delta f(\mathbf{X} \mid \mathbf{C})\, d\mathbf{X} \right) \Bigg] d\pi(\mathbf{C}).\end{aligned} \tag{2}$$

Simplify the second term:

$$\begin{aligned}\delta\mathcal{L} = \int \Bigg[&- \int_S \delta f(\mathbf{X} \mid \mathbf{C}) (\ln f(\mathbf{X} \mid \mathbf{C}) + 1)\, d\mathbf{X} + \lambda(\mathbf{C}) \left(- \int_S \delta f(\mathbf{X} \mid \mathbf{C})\, d\mathbf{X} \right) \\ &- \sum_{i=1}^{n} \mu_i(\mathbf{C}) \int_S \alpha_i(\mathbf{X}) \delta f(\mathbf{X} \mid \mathbf{C})\, d\mathbf{X} \Bigg] d\pi(\mathbf{C}).\end{aligned} \tag{3}$$

Combine the terms:

$$\delta\mathcal{L} = \int \left[\int_S \delta f(\mathbf{X} \mid \mathbf{C}) \left(-(\ln f(\mathbf{X} \mid \mathbf{C}) + 1) - \lambda(\mathbf{C}) - \sum_{i=1}^{n} \mu_i(\mathbf{C}) \alpha_i(\mathbf{X}) \right) d\mathbf{X} \right] d\pi(\mathbf{C}). \tag{4}$$

For $\delta\mathcal{L} = 0$, the integrand must be zero for all $\delta f(\mathbf{X} \mid \mathbf{C})$:

$$-(\ln f(\mathbf{X} \mid \mathbf{C}) + 1) - \lambda(\mathbf{C}) - \sum_{i=1}^{n} \mu_i(\mathbf{C}) \alpha_i(\mathbf{X}) = 0. \tag{5}$$

Solve this equation to get:

$$\ln f(\mathbf{X} \mid \mathbf{C}) = -1 - \lambda(\mathbf{C}) - \sum_{i=1}^{n} \mu_i(\mathbf{C}) \alpha_i(\mathbf{X}). \tag{6}$$

Exponentiate both sides:

$$f(\mathbf{X} \mid \mathbf{C}) = \exp\left(-1 - \lambda(\mathbf{C}) - \sum_{i=1}^{n} \mu_i(\mathbf{C}) \alpha_i(\mathbf{X}) \right). \tag{7}$$

Now apply the first constraint:

$$\int_S f(\mathbf{X} \mid \mathbf{C}) \, d\mathbf{X} = 1. \tag{8}$$

Substitute $f(\mathbf{X} \mid \mathbf{C})$:

$$\int_S \exp\left(-1 - \lambda(\mathbf{C}) - \sum_{i=1}^{n} \mu_i(\mathbf{C}) \alpha_i(\mathbf{X}) \right) d\mathbf{X} = 1. \tag{9}$$

Factor out the constant term:

$$\exp(-1 - \lambda(\mathbf{C})) \int_S \exp\left(-\sum_{i=1}^{n} \mu_i(\mathbf{C}) \alpha_i(\mathbf{X}) \right) d\mathbf{X} = 1. \tag{10}$$

Define:

$$Z(\mathbf{C}) = \int_S \exp\left(-\sum_{i=1}^{n} \mu_i(\mathbf{C}) \alpha_i(\mathbf{X}) \right) d\mathbf{X}, \tag{11}$$

Then:

$$\exp(-1 - \lambda(\mathbf{C})) \cdot Z(\mathbf{C}) = 1. \tag{12}$$

Solve for $\lambda(\mathbf{C})$:

$$\exp(-1 - \lambda(\mathbf{C})) = \frac{1}{Z(\mathbf{C})}. \tag{13}$$

Take the natural logarithm:

$$-1-\lambda(\mathbf{C})=\ln\left(\frac{1}{Z(\mathbf{C})}\right)=-\ln Z(\mathbf{C}), \tag{14}$$

Thus:

$$\lambda(\mathbf{C})=\ln Z(\mathbf{C})-1. \tag{15}$$

So, $f(\mathbf{X}\mid\mathbf{C})$ can be written as:

$$f(\mathbf{X}\mid\mathbf{C})=\exp\left(-1-(\ln Z(\mathbf{C})-1)-\sum_{i=1}^{n}\mu_i(\mathbf{C})\alpha_i(\mathbf{X})\right), \tag{16}$$

Or:

$$f(\mathbf{X}\mid\mathbf{C})=\frac{1}{Z(\mathbf{C})}\exp\left(-\sum_{i=1}^{n}\mu_i(\mathbf{C})\alpha_i(\mathbf{X})\right). \tag{17}$$

Now apply the second constraint:

$$\int_S \alpha_i(\mathbf{X})f(\mathbf{X}\mid\mathbf{C})\,d\mathbf{X}=\mathbf{g}_i(\mathbf{C}), \tag{18}$$

We have:

$$\int_S \alpha_i(\mathbf{X})\frac{1}{Z(\mathbf{C})}\exp\left(-\sum_{i=1}^{n}\mu_i(\mathbf{C})\alpha_i(\mathbf{X})\right)\,d\mathbf{X}=\mathbf{g}_i(\mathbf{C}). \tag{19}$$

So:

$$\frac{1}{Z(\mathbf{C})}\int_S \alpha_i(\mathbf{X})\exp\left(-\sum_{i=1}^{n}\mu_i(\mathbf{C})\alpha_i(\mathbf{X})\right)\,d\mathbf{X}=\mathbf{g}_i(\mathbf{C}). \tag{20}$$

Therefore, we find:

$$Z(\mathbf{C})=\int_S \exp\left(-\sum_{i=1}^{n}\mu_i(\mathbf{C})\alpha_i(\mathbf{X})\right)\,d\mathbf{X}. \tag{21}$$

This provides the relationship between $\lambda(\mathbf{C})$ and $\mu_i(\mathbf{C})$, and the form of $f(\mathbf{X}\mid\mathbf{C})$.

2 Derivation of the Partial Differential Equation

Given the equation:

$$\frac{1}{Z(\mathbf{C})}\int_S \alpha_i(\mathbf{X})\exp\left(-\sum_{j=1}^{n}\mu_j(\mathbf{C})\alpha_j(\mathbf{X})\right)\,d\mathbf{X}=E[\alpha_i(\mathbf{X})\mid\mathbf{C}], \tag{22}$$

we need to derive:

$$\frac{\partial \mu_i(\mathbf{C})}{\partial \mathbf{C}_j} = -\frac{\frac{\partial}{\partial \mathbf{C}_j} E[\alpha_i(\mathbf{X}) \mid \mathbf{C}]}{\mathrm{Var}(\alpha_i(\mathbf{X}) \mid \mathbf{C})}. \tag{23}$$

We assume further that the measurements α are orthogonal with each other. First, we express the expectation and variance:

$$E[\alpha_i(\mathbf{X}) \mid \mathbf{C}] = \frac{1}{Z(\mathbf{C})} \int_S \alpha_i(\mathbf{X}) \exp\left(-\sum_{k=1}^{n} \mu_k(\mathbf{C}) \alpha_k(\mathbf{X}) \right) d\mathbf{X}, \tag{24}$$

where

$$Z(\mathbf{C}) = \int_S \exp\left(-\sum_{k=1}^{n} \mu_k(\mathbf{C}) \alpha_k(\mathbf{X}) \right) d\mathbf{X}. \tag{25}$$

The conditional variance is given by:

$$\mathrm{Var}(\alpha_i(\mathbf{X}) \mid \mathbf{C}) = E[\alpha_i^2(\mathbf{X}) \mid \mathbf{C}] - (E[\alpha_i(\mathbf{X}) \mid \mathbf{C}])^2, \tag{26}$$

where

$$E[\alpha_i^2(\mathbf{X}) \mid \mathbf{C}] = \frac{1}{Z(\mathbf{C})} \int_S \alpha_i^2(\mathbf{X}) \exp\left(-\sum_{k=1}^{n} \mu_k(\mathbf{C}) \alpha_k(\mathbf{X}) \right) d\mathbf{X}. \tag{27}$$

Next, we find the partial derivative of the expectation with respect to $\mathbf{C}$:

$$\frac{\partial}{\partial C_j} E[\alpha_i(\mathbf{X}) \mid \mathbf{C}] = \frac{\partial}{\partial C_j} \left(\frac{1}{Z(\mathbf{C})} \int_S \alpha_i(\mathbf{X}) \exp\left(-\sum_{k=1}^{n} \mu_k(\mathbf{C}) \alpha_k(\mathbf{X}) \right) d\mathbf{X} \right). \tag{28}$$

Using the quotient rule, we get:

$$\frac{\partial}{\partial C_j} \left(\frac{1}{Z(\mathbf{C})} \right) = -\frac{\frac{\partial Z(\mathbf{C})}{\partial C_j}}{Z(\mathbf{C})^2}. \tag{29}$$

Thus,

$$\begin{aligned}\frac{\partial}{\partial C_j} E[\alpha_i(\mathbf{X}) \mid \mathbf{C}] = &-\frac{\frac{\partial Z(\mathbf{C})}{\partial C_j}}{Z(\mathbf{C})^2} \int_S \alpha_i(\mathbf{X}) \exp\left(-\sum_{k=1}^{n} \mu_k(\mathbf{C}) \alpha_k(\mathbf{X}) \right) d\mathbf{X} \\ &+ \frac{1}{Z(\mathbf{C})} \frac{\partial}{\partial C_j} \left(\int_S \alpha_i(\mathbf{X}) \exp\left(-\sum_{k=1}^{n} \mu_k(\mathbf{C}) \alpha_k(\mathbf{X}) \right) d\mathbf{X} \right).\end{aligned} \tag{30}$$

Using the chain rule:

$$
\begin{aligned}
&\frac{\partial}{\partial C_j}\left(\int_S \alpha_i(\mathbf{X}) \exp\left(-\sum_{k=1}^{n} \mu_k(\mathbf{C})\alpha_k(\mathbf{X})\right) d\mathbf{X}\right) \\
&\quad = \int_S \alpha_i(\mathbf{X}) \left(-\sum_{k=1}^{n} \frac{\partial \mu_k(\mathbf{C})}{\partial C_j}\alpha_k(\mathbf{X})\right) \cdot \exp\left(-\sum_{k=1}^{n} \mu_k(\mathbf{C})\alpha_k(\mathbf{X})\right) d\mathbf{X}.
\end{aligned} \tag{31}
$$

Using the orthogonality condition:

$$
\int_S \alpha_i(\mathbf{X})\alpha_j(\mathbf{X}) \exp\left(-\sum_{k=1}^{n} \mu_k(\mathbf{C})\alpha_k(\mathbf{X})\right) d\mathbf{X} = \delta_{ij} \int_S \alpha_i^2(\mathbf{X}) \exp\left(-\mu_i(\mathbf{C})\alpha_i(\mathbf{X})\right) d\mathbf{X}. \tag{32}
$$

Thus, we get:

$$
\begin{aligned}
\frac{\partial}{\partial C_j} E[\alpha_i(\mathbf{X}) \mid \mathbf{C}] &= -\frac{\frac{\partial Z(\mathbf{C})}{\partial C_j}}{Z(\mathbf{C})^2} I_i(\mathbf{C}) \\
&\quad - \frac{1}{Z(\mathbf{C})} \sum_{k=1}^{n} \frac{\partial \mu_k(\mathbf{C})}{\partial C_j} \int_S \alpha_i(\mathbf{X})\alpha_k(\mathbf{X}) \exp\left(-\sum_{k=1}^{n} \mu_k(\mathbf{C})\alpha_k(\mathbf{X})\right) d\mathbf{X}.
\end{aligned} \tag{33}
$$

Given the orthogonality of $\alpha_i(\mathbf{X})$ and $\alpha_j(\mathbf{X})$, we simplify the above into the following:

$$
\frac{\partial}{\partial C_j} E[\alpha_i(\mathbf{X}) \mid \mathbf{C}] = -\frac{\frac{\partial Z(\mathbf{C})}{\partial C_j}}{Z(\mathbf{C})^2} I_i(\mathbf{C}) - \frac{1}{Z(\mathbf{C})} \frac{\partial \mu_i(\mathbf{C})}{\partial C_j} \int_S \alpha_i^2(\mathbf{X}) \exp\left(-\mu_i(\mathbf{C})\alpha_i(\mathbf{X})\right) d\mathbf{X}. \tag{34}
$$

The variance is:

$$
\operatorname{Var}(\alpha_i(\mathbf{X}) \mid \mathbf{C}) = \frac{1}{Z(\mathbf{C})} \int_S \alpha_i^2(\mathbf{X}) \exp\left(-\mu_i(\mathbf{C})\alpha_i(\mathbf{X})\right) d\mathbf{X} - \left(\frac{I_i(\mathbf{C})}{Z(\mathbf{C})}\right)^2. \tag{35}
$$

Therefore, the partial differential equation simplifies to:

$$
\frac{\partial \mu_i(\mathbf{C})}{\partial \mathbf{C}_j} = -\frac{\frac{\partial}{\partial \mathbf{C}_j} E[\alpha_i(\mathbf{X}) \mid \mathbf{C}]}{\operatorname{Var}(\alpha_i(\mathbf{X}) \mid \mathbf{C})}. \tag{36}
$$

And the conditional probability density function is given as:

$$
f(\mathbf{X} \mid \mathbf{C}) = \frac{1}{Z(\mathbf{C})} \exp\left(-\sum_{i=1}^{n} \mu_i(\mathbf{C})\alpha_i(\mathbf{X})\right), \tag{37}
$$

which is a member of exponential family, where

$$Z(\mathbf{C}) = \int_S \exp\left(-\sum_{k=1}^{n} \mu_k(\mathbf{C})\alpha_k(\mathbf{X})\right)\, d\mathbf{X}. \tag{38}$$