Natural and Exponential Dispersion Families

Søren Lund Pedersen

In this article, I would like to discuss the motivation, intuition, and the components that characterize a distribution as a member of the Natural Dispersion Family (NDF) and its extension to Exponential Dispersion Families (EDF).

The Motivation

The main motivation behind Natural Dispersion Families is to summarize a variety of well-known distributions into a concise and structured format. By representing distributions within the NDF framework, we harness properties that streamline analysis and modeling. The transformation of a distribution into an NDF form offers several advantages:

- It facilitates the use of generalized linear models (GLMs), which are foundational for methods such as logistic regression and zero-inflated models.
- It allows for the application of conjugate priors in Bayesian statistics, greatly simplifying the process of updating beliefs and interpreting results.

The NDF framework encapsulates distributions in a standardized way, paving the path for more efficient computation and clearer theoretical understanding.

The Intuition

Understanding Natural Exponential Families (NEF) and Exponential Dispersion Families (EDF) requires recognizing the relationship between variance and mean. In a basic NEF, the variance is directly linked to the mean—often proportional to it. However, real-world data sometimes exhibit variance patterns that cannot be captured by a simple mean-variance relationship. EDF generalizes NEF by introducing an additional dispersion parameter, allowing models to accommodate situations where variance is not a simple function of the mean.

An NDF expresses the probability distribution in an exponential form:

$$f(x \mid \theta) = h(x) \exp\big(\theta\, T(x) - A(\theta)\big),$$

where:

- θ is the canonical (natural) parameter,
- $T(x)$ is the sufficient statistic,
- $h(x)$ is a base measure independent of θ,
- $A(\theta)$ is the log-partition function ensuring proper normalization.

This form implies that all information about the data relevant to estimating θ is captured by the sufficient statistic $T(x)$. For instance, in the Poisson distribution, the sum $\sum_i x_i$ acts as a sufficient statistic, summarizing all necessary information from the data.

A distribution qualifies as an NDF member if its variance is tied to the mean through a specific variance function. The selection of parameters is done such that the exponent's dependency on the data remains linear, which greatly simplifies analysis and computation.

The advantages of using NDFs include:

- **Computational Efficiency**: Focusing on sufficient statistics reduces data storage needs and computational overhead. For example, storing the sum $\sum_i X_i$ instead of every individual x_i can lead to significant savings.
- **Simplified Differentiation**: The linearity of parameters in the exponent makes derivation of estimators, like the Maximum Likelihood Estimator (MLE), more straightforward.

1 General Form of the Density

A single-parameter NEF can be written as:

$$f(x \mid \theta) = h(x) \exp\big(\theta\, T(x) - A(\theta)\big),$$

where:

- θ is the canonical (natural) parameter,
- $T(x)$ is the sufficient statistic for θ,
- $h(x)$ is a function that does not depend on θ,
- $A(\theta)$ is the log-partition (cumulant) function that ensures the distribution integrates (or sums) to 1.

For the multi-parameter case, where both $\boldsymbol{\theta}$ and $\mathbf{T}(x)$ are vectors, we have:

$$f(x \mid \boldsymbol{\theta}) = h(x) \exp\Big(\boldsymbol{\theta}^{\mathsf{T}}\, \mathbf{T}(x) - A(\boldsymbol{\theta})\Big).$$

The concepts of the **Canonical Parameter** and **Sufficient Statistics** are central to this formulation, so let us delve deeper into those.

2 Canonical Parameter

The Canonical Parameter provides a standardized way to express a probability distribution. To find the canonical parameter for a given distribution, one must rewrite the distribution in the NEF form. This involves:

- Identifying the natural parameter of the distribution.
- Matching the terms of the original distribution to the NEF form.

For example, consider a fair six-sided die. Each outcome has an equal probability of $p = \frac{1}{6}$. The canonical parameter in this case relates to the log-probabilities of outcomes. By converting these probabilities into their logarithmic form, we express the distribution in a standardized exponential family form, making it amenable to analysis using NEF tools.

3 Sufficient Statistics

Sufficient statistics summarize all the necessary information in the data relevant to estimating a parameter, rendering the rest of the data redundant for that purpose. Once the sufficient statistic is determined, the raw data no longer provide additional insight for estimating the parameter.

For example, to determine if a die is fair, one does not need to retain every roll outcome. Instead, one can summarize the outcomes by counting the frequency of each face. If these frequencies are approximately equal (i.e., each face appears roughly $\frac{1}{6}$ of the time), one can conclude with high probability that the die is fair.

4 Log-partition

The **log-partition function** $A(\theta)$ ensures that the probability distribution is properly normalized (sums/integrates to 1). It is defined by the condition that the total probability is unity:

$$\int f(x \mid \theta)dx = 1 \quad \text{for continuous distributions}$$

$$\sum f(x \mid \theta) = 1 \quad \text{for discrete distributions}$$

The log-partition function can be computed as:

$$A(\theta) = \log \int h(x) \exp(\theta T(x))dx$$

for continuous distributions, or

$$A(\theta) = \log \sum_x h(x) \exp(\theta T(x))$$

for discrete distributions.

5 Example: Poisson Distribution in NDF

Consider a Poisson-distributed random variable X with parameter λ:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots$$

To express this distribution in the Natural Exponential Family (NEF) form, we proceed with the following steps:

1. **Rewrite the Poisson pmf:**

$$P(X = x) = \frac{1}{x!} \exp\bigl(x \log \lambda - \lambda\bigr).$$

2. **Identify components to match the NEF form:** The NEF form is given by

$$f(x \mid \theta) = h(x) \exp\Bigl(\theta\, T(x) - A(\theta)\Bigr).$$

Comparing this with the Poisson form:

$$h(x) = \frac{1}{x!}, \quad T(x) = x.$$

3. **Determine the canonical parameter θ:** We set

$$\theta = \log \lambda,$$

   so that

$$\lambda = e^{\theta}.$$

4. **Identify the log-partition function $A(\theta)$:** From the rewritten pmf, we have the term $-\lambda = -e^{\theta}$ in the exponent. Thus,

$$A(\theta) = e^{\theta}.$$

5. **Assemble the NEF representation:** Substituting back, the Poisson pmf can be written as

$$P(X = x) = \underbrace{\frac{1}{x!}}_{h(x)} \exp\Big(x\theta - e^{\theta}\Big),$$

   which matches the NEF form with the identified components:

$$h(x) = \frac{1}{x!}, \quad T(x) = x, \quad \theta = \log \lambda, \quad A(\theta) = e^{\theta}.$$

This example illustrates how the Poisson distribution fits into the Natural Exponential Family framework by identifying the sufficient statistic $T(x) = x$, the canonical parameter $\theta = \log \lambda$, and the log-partition function $A(\theta) = e^{\theta}$.

6 Exponential Dispersion Families

While NDFs are powerful, they constrain the variance to be a function of the mean. This limitation can be restrictive when modeling data where variance and mean vary independently, such as in the normal distribution $N(\mu, \sigma^2)$.

The Exponential Dispersion Family (EDF) extends the NDF to incorporate an additional dispersion parameter ϕ. The EDF is formally written as:

$$f(y \mid \theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\right).$$

This resembles the NDF form, with the following components:

- θ is the canonical (natural) parameter,
- $b(\theta)$ is the log normalizer, analogous to $A(\theta)$ in NDF,
- ϕ is the dispersion parameter that controls the variance independent of the mean,
- $c(y, \phi)$ is a term that does not depend on θ.

In some distributions like the Poisson and binomial, the dispersion parameter ϕ is fixed at 1, which reduces the EDF to the familiar NDF/NEF form. However, when data exhibit overdispersion or underdispersion (variance greater or less than the mean), the EDF allows for modeling these phenomena by adjusting ϕ.

7 Canonical Link Function

Within the framework of generalized linear models (GLMs), the **canonical link function** plays a pivotal role. It links the mean μ of the response variable to the linear predictor η through the natural parameter θ and the log normalizer derivative:

$$g(\mu) = b'^{-1}(\mu).$$

Since $\mu = b'(\theta)$, the canonical link function satisfies $g(\mu) = \theta$, establishing a direct relationship between the mean response and the canonical parameter.

For instance:

- In the binomial distribution, the canonical link function is the logit function $\log \frac{\mu}{1-\mu}$.
- In the Poisson distribution, the canonical link is the log function $\log(\mu)$.

This direct relationship simplifies parameter estimation and model interpretation. Additionally, for EDFs:

$$b'(\theta) = \mu$$

provides the mean, and

$$\phi\, b''(\theta) = \mathrm{Var}(X)$$

gives the variance, showing how the second derivative of the log normalizer scaled by ϕ dictates variability.

8 Example: Normal Distribution in EDF

Consider a normally distributed random variable $Y \sim N(\mu, \sigma^2)$ with probability density function:

$$f(y \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(y-\mu)^2}{2\sigma^2}\Big).$$

We aim to express this density in the EDF form:

$$f(y \mid \theta, \phi) = \exp\Big(\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\Big).$$

1. **Rewrite the Normal density:**

   Expand the quadratic term:

   $$-\frac{(y-\mu)^2}{2\sigma^2} = -\frac{y^2 - 2\mu y + \mu^2}{2\sigma^2} = -\frac{y^2}{2\sigma^2} + \frac{\mu y}{\sigma^2} - \frac{\mu^2}{2\sigma^2}.$$

2. **Express the density in exponential form:**

   $$f(y \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(\frac{\mu y - \mu^2/2}{\sigma^2} - \frac{y^2}{2\sigma^2}\Big).$$

3. **Identify EDF components:**

   Set

   $$\phi = \sigma^2,$$

and observe that the exponent contains

$$\frac{\mu y - \mu^2/2}{\sigma^2} = \frac{y\theta - \theta^2/2}{\phi}$$

with $\theta = \mu$. The remaining terms combine into a function of y and ϕ only.

4. **Determine the log normalizer $b(\theta)$:**

   From the term $-\theta^2/2$, we identify

$$b(\theta) = \frac{\theta^2}{2}.$$

5. **Identify the remaining term $c(y, \phi)$:**

   The terms independent of θ are grouped as

$$c(y, \phi) = -\frac{y^2}{2\phi} - \frac{1}{2}\log(2\pi\phi).$$

6. **Assemble the EDF representation:**

   Substituting these components back, the Normal density can be written as

$$f(y \mid \theta, \phi) = \exp\Big(\frac{y\theta - \theta^2/2}{\phi} + c(y, \phi)\Big),$$

   where

$$\theta = \mu, \quad \phi = \sigma^2, \quad b(\theta) = \frac{\theta^2}{2}, \quad c(y, \phi) = -\frac{y^2}{2\phi} - \frac{1}{2}\log(2\pi\phi).$$

This example demonstrates how the Normal distribution fits into the Exponential Dispersion Family by identifying its canonical parameter $\theta = \mu$, dispersion parameter $\phi = \sigma^2$, log normalizer $b(\theta) = \frac{\theta^2}{2}$, and the function $c(y, \phi)$.