

# Mathematical Statistics

## Tutorial 1

1. Prove that the Multinomial distribution  $\mathbf{X} \sim \text{Multinomial}(n, \mathbf{p})$  is a  **$k - 1$  parameter exponential family**.
2. Let  $X$  be a random variable with Probability Density Function (PDF) given by

$$f(x | \theta) = (1 - \theta)e^x \cdot \mathbf{I}_{\{x < 0\}} + \theta^2 e^{-\theta x} \cdot \mathbf{I}_{\{x \geq 0\}}, \quad \text{where } \theta \in (0, 1).$$

Prove that this distribution is a 2-parameter exponential family.

## Excercise 1

Prove that the Multinomial distribution  $\mathbf{X} \sim \text{Multinomial}(n, \mathbf{p})$  is a  **$k - 1$  parameter exponential family**. The **Probability Mass Function (PMF)** for  $\mathbf{X} \sim \text{Multinomial}(n, \mathbf{p})$  is:

$$f(\mathbf{x}; n, \mathbf{p}) = \binom{n}{\mathbf{x}} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}, \quad \text{where } \binom{n}{\mathbf{x}} = \frac{n!}{x_1! x_2! \cdots x_k!}.$$

The Multinomial distribution is the generalization of the Binomial distribution. It models the counts resulting from a fixed number of independent trials ( $n$ ), where each trial can result in one of  $k$  distinct categories.

- **Trials ( $n$ ):** Fixed number of independent trials.
  - **Probabilities ( $\mathbf{p}$ ):** A vector  $\mathbf{p} = (p_1, \dots, p_k)$  where  $p_j$  is the probability of outcome  $j$ , constrained by  $\sum_{j=1}^k p_j = 1$ .
  - **Variable ( $\mathbf{X}$ ):** A random vector of counts  $\mathbf{X} = (X_1, \dots, X_k)$ , where  $X_j$  is the count for category  $j$ , constrained by  $\sum_{j=1}^k X_j = n$ .
- 

### Tasks:

1. **Transform** the Multinomial PMF,  $f(\mathbf{x}; n, \mathbf{p})$ , into the canonical form of a  $k - 1$  parameter exponential family:

$$f(\mathbf{x}; \boldsymbol{\eta}) = h(\mathbf{x}) \exp \left\{ \sum_{j=1}^{k-1} \eta_j T_j(\mathbf{x}) - A(\boldsymbol{\eta}) \right\}.$$

2. **Explicitly identify** the following components of the canonical form based on the original parameters ( $\mathbf{p}$ ) and variables ( $\mathbf{x}$ ):

- The base measure  $h(\mathbf{x})$ .
- The vector of statistics  $\mathbf{T}(\mathbf{x}) = (T_1(\mathbf{x}), \dots, T_{k-1}(\mathbf{x}))^\top$ .
- The vector of natural parameters  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_{k-1})^\top$ .
- The log-normalizer function  $A(\boldsymbol{\eta})$ , expressed as a function of  $\boldsymbol{\eta}$ .

Using the  $k$ -th category as the reference ( $p_k$  and  $X_k$ ), the Probability Mass Function (PMF) in the canonical form is:

$$\begin{aligned} f(\mathbf{x}; \boldsymbol{\eta}) &= \binom{n}{\mathbf{x}} \exp \left\{ \sum_{j=1}^{k-1} \log(p_j) x_j + x_k \log(p_k) \right\} \\ &= \binom{n}{\mathbf{x}} \exp \left\{ \sum_{j=1}^{k-1} \log(p_j) x_j + (n - \sum_{j=1}^{k-1} x_j) \log(p_k) \right\} \\ &= \underbrace{\binom{n}{\mathbf{x}}}_{h(\mathbf{x})} \exp \left\{ \sum_{j=1}^{k-1} \underbrace{\log\left(\frac{p_j}{p_k}\right)}_{\eta_j} \underbrace{x_j}_{T_j(\mathbf{x})} - \underbrace{[-n \log(p_k)]}_{A(\boldsymbol{\eta})} \right\} \end{aligned}$$

with  $\boldsymbol{\eta} = (n, \mathbf{p})$

This is due to the inherent constraint  $\sum p_j = 1$ , meaning that only  $k - 1$  parameters are functionally independent.

1. **Vector of Statistics  $\mathbf{T}(\mathbf{x})$ :** The sufficient statistic is the vector of the first  $k - 1$  counts:

$$\mathbf{T}(\mathbf{x}) = (X_1, X_2, \dots, X_{k-1})^\top.$$

2. **Vector of Natural Parameters  $\boldsymbol{\eta}$ :**

$$\boldsymbol{\eta} = \left( \log \left( \frac{p_1}{p_k} \right), \dots, \log \left( \frac{p_{k-1}}{p_k} \right) \right)^\top.$$

3. **Log-Normalizer Function  $A(\boldsymbol{\eta})$ :** Expressing  $p_k$  in terms of  $\boldsymbol{\eta}$  allows  $A(\boldsymbol{\eta})$  to be written solely as a function of the natural parameters:

$$A(\boldsymbol{\eta}) = -n \log(p_k) = n \log \left( 1 + \sum_{j=1}^{k-1} e^{\eta_j} \right).$$

since  $e^{\eta_j} = \frac{p_j}{p_k}$  and  $\sum_{j=1}^{k-1} e^{\eta_j} = \frac{1-p_k}{p_k}$ , therefore  $1 + \sum_{j=1}^{k-1} e^{\eta_j} = \frac{1}{p_k}$  and  $\log(1 + \sum_{j=1}^{k-1} e^{\eta_j}) = -\log p_k$

Note that the Multinomial distribution can be written using  $k$  exponential terms. The PMF is written by incorporating all  $p_j$ 's into the exponential term:

$$f(\mathbf{x} \mid \boldsymbol{\eta}) = \underbrace{\binom{n}{\mathbf{x}}}_{h(\mathbf{x})} \exp \left\{ \sum_{j=1}^k \underbrace{\log(p_j)}_{\eta_j} \underbrace{x_j}_{T_j(\mathbf{x})} \right\}.$$

1. **Vector of Statistics  $\mathbf{T}(\mathbf{x})$ :** The vector includes all  $k$  counts:

$$\mathbf{T}(\mathbf{x}) = (X_1, X_2, \dots, X_k)^\top.$$

2. **Vector of Natural Parameters  $\boldsymbol{\eta}$ :**

$$\boldsymbol{\eta} = (\log(p_1), \log(p_2), \dots, \log(p_k))^\top.$$

This  $k$ -parameter form is correct but the  $k - 1$  form is preferred for theoretical statistical analysis.

## Excercise 2

Let  $X$  be a random variable wit Probability Density Function (PDF) given by

$$f(x \mid \theta) = (1 - \theta)e^x \cdot \mathbf{I}_{\{x < 0\}} + \theta^2 e^{-\theta x} \cdot \mathbf{I}_{\{x \geq 0\}}, \quad \text{where } \theta \in (0, 1).$$

Prove that this distribution is a 2- parameter exponential family.

We aim to write  $f(x \mid \theta)$  in the canonical form:

$$f(x \mid \boldsymbol{\eta}) = h(x) \exp \{ \eta_1 T_1(x) + \eta_2 T_2(x) - A(\boldsymbol{\eta}) \}.$$

Note that

$$f(x \mid \theta) = [(1 - \theta)e^x]^{\mathbf{I}_{\{x < 0\}}} [\theta^2 e^{-\theta x}]^{\mathbf{I}_{\{x \geq 0\}}}$$

We use  $\mathbf{I}_{\{x<0\}} = 1 - \mathbf{I}_{\{x\geq 0\}}$  and we have that

$$\begin{aligned}
f(x \mid \theta) &= [(1 - \theta)e^x]^{1 - \mathbf{I}_{\{x\geq 0\}}} [\theta^2 e^{-\theta x}]^{\mathbf{I}_{\{x\geq 0\}}} \\
&= (1 - \theta)e^x [(1 - \theta)^{-1} e^{-x} \theta^2 e^{-\theta x}]^{\mathbf{I}_{\{x\geq 0\}}} \\
&= (1 - \theta)e^{x - x\mathbf{I}_{\{x\geq 0\}}} e^{2\log(\theta/1-\theta)\mathbf{I}_{\{x\geq 0\}} - \theta x\mathbf{I}_{\{x\geq 0\}}} \\
&= e^{x - x\mathbf{I}_{\{x\geq 0\}}} e^{2\log(\theta/1-\theta)\mathbf{I}_{\{x\geq 0\}} - \theta x\mathbf{I}_{\{x\geq 0\}} - \log(1-\theta)}
\end{aligned}$$

$$f(x \mid \theta) = \underbrace{e^{x - x\mathbf{I}_{\{x\geq 0\}}}}_{h(x)} e^{\underbrace{2\log(\theta/1-\theta)\mathbf{I}_{\{x\geq 0\}}}_{\eta_1} + \underbrace{-\theta}_{\eta_2} \underbrace{x\mathbf{I}_{\{x\geq 0\}}}_{T_2(x)} - \underbrace{\log(1-\theta)}_{A(\theta)}}$$

1.  $h(x)$ :

$$h(x) = \exp(x - x \cdot \mathbf{I}_{\{x\geq 0\}}) = \exp(x \cdot \mathbf{I}_{\{x<0\}})$$

2. **Statistics  $\mathbf{T}(x)$ :**

$$\mathbf{T}(x) = \begin{pmatrix} T_1(x) \\ T_2(x) \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{\{x\geq 0\}} \\ x\mathbf{I}_{\{x\geq 0\}} \end{pmatrix}$$

3. **Natural Parameters  $\boldsymbol{\eta}$ :**

$$\boldsymbol{\eta} = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} 2\log\left(\frac{\theta}{1-\theta}\right) \\ -\theta \end{pmatrix}$$

4.  $A(\boldsymbol{\eta})$ :

$$A(\boldsymbol{\eta}) = \log(1 - \theta)$$