Exponential Family of Distributions

Aleksandar Dogandžić

January 26, 2017

Contents

READING: (Hero 2015, §3.5 and 4.4.1).

THE exponential family of distributions has the following form:[1]

$$\begin{aligned} f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x}\,|\,\boldsymbol{\theta}) &= h(\boldsymbol{x}) \exp\Big[\sum_{i=1}^{d} \eta_i(\boldsymbol{\theta}) T_i(\boldsymbol{x}) - B(\boldsymbol{\theta})\Big] \\ &= h(\boldsymbol{x}) \exp\big[\boldsymbol{\eta}^T(\boldsymbol{\theta}) \boldsymbol{T}(\boldsymbol{x}) - B(\boldsymbol{\theta})\big] \end{aligned} \tag{1}$$

where $\boldsymbol{\theta} \subset \mathbb{R}^d$ is a d-dimensional vector. By the factorization theorem, $\boldsymbol{T}(\boldsymbol{X}) = [T_1(\boldsymbol{X}), \ldots, T_d(\boldsymbol{X})]^T$ is sufficient for $\boldsymbol{\theta}$. It is the vector of *natural sufficient statistics* of the family. The exponential family is important: It covers quite a few useful distributions, including some that are fairly complex.[2]

Recall that the support of $f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x}\,|\,\boldsymbol{\theta})$ is the set of $\boldsymbol{x}$ for which the probability density function (pdf) is positive:

$$\operatorname{supp}\{f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x}\,|\,\boldsymbol{\theta})\} \triangleq \{\boldsymbol{x} \mid f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x}\,|\,\boldsymbol{\theta}) > 0\}.$$

✻ USEFUL tests:[3]

- if the support $\operatorname{supp}\{f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x}\,|\,\boldsymbol{\theta})\}$ of $f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x}\,|\,\boldsymbol{\theta})$ depends on $\boldsymbol{\theta}$, then $f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x}\,|\,\boldsymbol{\theta})$ *does not* belong to the exponential family of distributions. For example, uniform $U(0, \theta)$ is not a member of the exponential family.

- For pdfs and pmfs in the exponential family, the size d of $\boldsymbol{\theta}$ is the *same* as the dimension of the vector of natural sufficient statistics.

- Another simple way is to differentiate $\ln f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x}\,|\,\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{x}$ and examine the result. For example, for scalar θ and x, a pdf in the one-parameter exponential family (i.e., $d = 1$) is

$$f_{X|\Theta}(x\,|\,\theta) = h(x) \exp\big[\eta(\theta) T(x) - B(\theta)\big] \tag{2}$$

  and $\dfrac{\partial^2 \ln f_{X|\Theta}(x\,|\,\theta)}{\partial\theta\,\partial x}$ has separable form $\dfrac{\mathrm{d}\eta(\theta)}{\mathrm{d}\theta}\dfrac{\mathrm{d}T(x)}{\mathrm{d}x}$. Similar tests can be constructed for pmfs or parameter vectors $\boldsymbol{\theta}$ or vector data $\boldsymbol{x}$.

[1] The probability mass function (pmf) form of the exponential family is analogous. See (Hero 2015, § 3.5.4 and § 3.5.5) for more about the exponential family of distributions.

[2] For example, Markov random fields, used in image analysis, are typically in the exponential-family form.

[3] Read also (Hero 2015, §3.5.5).

⁎ INDEPENDENT, IDENTICALLY DISTRIBUTED (I.I.D.) measurements. Suppose that we have multiple i.i.d. measurements $\boldsymbol{X} = (X[n])_{n=0}^{N-1}$ coming from the above pdf. Then,

$$f_{\boldsymbol{X}\,|\,\boldsymbol{\Theta}}(\boldsymbol{x}\,|\,\theta) = \Big[\prod_{n=0}^{N-1} h(x[n])\Big] \exp\{\boldsymbol{\eta}^T(\boldsymbol{\theta})\boldsymbol{T}(x[n]) - B(\boldsymbol{\theta})\}$$

$$= \underbrace{\Big[\prod_{n=0}^{N-1} h(x[n])\Big] \exp\Big\{\boldsymbol{\eta}^T(\boldsymbol{\theta})\Big[\sum_{n=0}^{N-1} \boldsymbol{T}(x[n])\Big] - NB(\boldsymbol{\theta})\Big\}}_{\text{exponential family}}$$

and, hence, the vector of natural sufficient statistics is

$$\boldsymbol{T}(\boldsymbol{x}) = \sum_{n=0}^{N-1} \boldsymbol{T}(x[n])$$

which is also minimal (Bickel and Doksum 2001).

Transform invariance and popular parameterizations

TRANSFORMATION of the parameters $\boldsymbol{\theta}$ by an invertible function preserves membership in the exponential family.

⁎ EXAMPLE. Consider $d = 1$, i.e.,

$$f_{X\,|\,\Theta}(x\,|\,\theta) = h(x)\exp\big[\eta(\theta)T(x) - B(\theta)\big]$$

and $\alpha = 1/\theta$. Then,

$$f_{X\,|\,\alpha}(x\,|\,\alpha) = h(x)\exp\big[\underbrace{\eta(1/\alpha)}_{\eta_{\text{new}}(\alpha)}\,T(x) - \underbrace{B(1/\alpha)}_{B_{\text{new}}(\alpha)}\big].$$

⁎ CANONICAL form with natural parametrization. Pick the following invertible transform

$$\boldsymbol{\eta} = \boldsymbol{\eta}(\boldsymbol{\theta})$$

$\boldsymbol{\eta}$ is the *canonical parameter vector.*

with inverse $\boldsymbol{\theta}(\boldsymbol{\eta})$. Then

$$f_{\boldsymbol{X}\,|\,\boldsymbol{\eta}}(\boldsymbol{x}\,|\,\boldsymbol{\eta}) = h(\boldsymbol{x})\exp\big[\boldsymbol{\eta}^T\boldsymbol{T}(\boldsymbol{x}) - A(\boldsymbol{\eta})\big]$$

canonical form with natural parametrization $\boldsymbol{\eta}$

where $A(\boldsymbol{\eta}) = B\big(\boldsymbol{\theta}(\boldsymbol{\eta})\big)$. In this case

$$\mathrm{E}_{\boldsymbol{X}\,|\,\boldsymbol{\eta}}\big[\boldsymbol{T}(\boldsymbol{X})\,|\,\boldsymbol{\eta}\big] = \frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \tag{3a}$$

$$\mathrm{cov}_{\boldsymbol{X}\,|\,\boldsymbol{\eta}}\big(\boldsymbol{T}(\boldsymbol{X})\,|\,\boldsymbol{\eta}\big) = \frac{\partial^2 A(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}\,\partial\boldsymbol{\eta}^T}. \tag{3b}$$

(3) is easy to show—we will see in handout `crb` that expected score is zero. See also (Bickel and Doksum 2001).

✻ MEAN-VALUE parameterization. Choose a parameterization $\boldsymbol{\theta}$ so that

$$\mathrm{E}_{\boldsymbol{X}\,|\,\boldsymbol{\theta}}[\boldsymbol{T}(\boldsymbol{X})\,|\,\boldsymbol{\theta}] = \boldsymbol{\theta} \tag{4}$$

i.e., the natural sufficient statistic $\boldsymbol{T}(\boldsymbol{X})$ is an unbiased estimator of the parameter vector $\boldsymbol{\theta}$.

☞ WE will see that $\boldsymbol{T}(\boldsymbol{X})$ is a minimum-variance unbiased (MVU) estimator of $\boldsymbol{\theta}$ in (4). Thus, mean-value parameterizations are special because the statement that an estimator is MVU is strong.

Examples

WE present several examples.

✻ EXPONENTIAL measurement.

$$\begin{aligned} f_{X\,|\,\Theta}(x\,|\,\theta) &= \mathrm{Expon}(x\,|\,\theta) \\ &= \theta \mathrm{e}^{-\theta x}\, \mathbb{1}_{[0,\infty)}(x) \qquad\qquad \theta > 0 \\ &= \mathbb{1}_{[0,\infty)}(x) \exp\left(-\theta x + \ln\theta\right) \end{aligned}$$

belongs to the one-parameter exponential family where $h(x) = \mathbb{1}_{[0,\infty)}(x)$, $T(x) = x$, $\eta(\theta) = -\theta$, $B(\theta) = -\ln\theta$.

✻ POISSON measurement.

$$\begin{aligned} f_{X\,|\,\Lambda}(x\,|\,\lambda) &= \mathrm{Poisson}(x\,|\,\lambda) \\ &= \frac{\lambda^x}{x!}\mathrm{e}^{-\lambda}\, \mathbb{1}_{\mathbb{Z}_{\geq 0}}(x) \\ &= \frac{\mathbb{1}_{\mathbb{Z}_{\geq 0}}(x)}{x!} \exp(x\ln\lambda - \lambda) \end{aligned}$$

with parameter $\lambda > 0$ belongs to the one-parameter exponential family:

$$f_{X\,|\,\Lambda}(x\,|\,\lambda) = h(x)\exp[\eta(\lambda)T(x) - B(\lambda)]$$

where $h(x) = \mathbb{1}_{\mathbb{Z}_{\geq 0}}(x)/x!$, $T(x) = x$, $\eta(\lambda) = \ln\lambda$, $B(\lambda) = \lambda$. Here, $\mathbb{Z}_{\geq 0}$ is the set of nonnegative integers.

✻ GAUSSIAN measurement with unknown mean and variance.

$$\begin{aligned} f_{\boldsymbol{X}\,|\,\boldsymbol{\Theta}}(\boldsymbol{x}\,|\,\boldsymbol{\theta}) &= \prod_{n=0}^{N-1} \mathcal{N}(x[n]\,|\,\mu, \sigma^2) \\ &= \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x[n]-\mu)^2}{2\sigma^2}\right\} \\ &= \exp\left\{-\frac{1}{2\sigma^2}\Big(\sum_{n=0}^{N-1} x^2[n]\Big) + \frac{\mu}{\sigma^2}\sum_{n=0}^{N-1} x[n]\right\} \frac{\exp(-0.5N\mu^2/\sigma^2)}{\sqrt{(2\pi\sigma^2)^N}} \end{aligned}$$

with parameters

$$\boldsymbol{\theta} = (\mu, \sigma^2)$$

belongs to the two-parameter exponential family:

$$f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x}\,|\,\boldsymbol{\theta}) = h(\boldsymbol{x}) \exp\left[\eta_1(\boldsymbol{\theta}) T_1(\boldsymbol{x}) + \eta_2(\boldsymbol{\theta}) T_2(\boldsymbol{x}) - B(\boldsymbol{\theta})\right]$$

where $h(\boldsymbol{x}) = 1, T_1(\boldsymbol{x}) = \sum_{n=0}^{N-1} x^2[n], T_2(\boldsymbol{x}) = \sum_{n=0}^{N-1} x[n], \eta_1(\boldsymbol{\theta}) = -1/(2\sigma^2), \eta_2(\boldsymbol{\theta}) = \mu/\sigma^2, B(\boldsymbol{\theta}) = 0.5N\mu^2/\sigma^2 + 0.5N\ln(2\pi\sigma^2)$.

✻ ONE-PARAMETER canonical exponential family. The one-parameter canonical exponential family with $d = 1$:

$$f_{\boldsymbol{X}|\eta}(\boldsymbol{x}|\eta) = h(\boldsymbol{x}) \exp\big[\underbrace{\eta}_{\substack{\text{scalar}\\\text{canonical}\\\text{parameter}}} T(\boldsymbol{x}) - A(\eta)\big]$$

where

$$A(\eta) = \ln \int \cdots \int h(\boldsymbol{\chi}) \exp\left[\eta T(\boldsymbol{\chi})\right] \mathrm{d}\boldsymbol{\chi} \qquad \text{for a pdf } f_{X|\eta}(\boldsymbol{x}\,|\,\eta)$$

$$A(\eta) = \ln \sum_{\boldsymbol{\chi}} h(\boldsymbol{\chi}) \exp\left[\eta T(\boldsymbol{\chi})\right]. \qquad \text{for a pmf } p_{X|\eta}(\boldsymbol{x}\,|\,\eta)$$

If we can compute the normalizing term $A(\eta)$ in a simple form, then we can find easily the mean and variance of $T(X)$ given η:

$$\mathrm{E}_{\boldsymbol{X}|\eta}[T(\boldsymbol{X})\,|\,\eta] = \frac{\mathrm{d}A(\eta)}{\mathrm{d}\eta}, \qquad \mathrm{var}_{\boldsymbol{X}|\eta}(T(\boldsymbol{X})\,|\,\eta) = \frac{\mathrm{d}^2 A(\eta)}{\mathrm{d}\eta^2} \tag{5}$$

which follows from (3).

Why is this useful?

✻ EXAMPLE. Suppose that $(X[n])_{n=0}^{N-1}$ are conditionally i.i.d. given θ, from

$$f_{X|\Theta}(x|\theta) = \frac{x}{\theta} \exp\left(-\frac{x^2}{2\theta}\right) \mathbb{1}_{[0,\infty)}(x)$$

Rayleigh pdf

where $\theta > 0$ is a parameter. Define $\boldsymbol{X} = (X[n])_{n=0}^{N-1}$ and $\boldsymbol{x} = (x[n])_{n=0}^{N-1}$ and write the model pdf as

$$\begin{aligned} f_{\boldsymbol{X}|\Theta}(\boldsymbol{x}\,|\,\theta) &= \prod_{n=0}^{N-1} \frac{x[n]}{\theta} \exp\left(-\frac{x^2[n]}{2\theta}\right) \\ &= \left(\prod_{n=0}^{N-1} x[n]\right) \exp\Bigg[\underbrace{-\frac{1}{2\theta}}_{\eta(\theta)} \underbrace{\left(\sum_{n=0}^{N-1} x^2[n]\right)}_{T(\boldsymbol{x})} - \underbrace{N\ln\theta}_{B(\theta)}\Bigg] \end{aligned}$$

implying

$$\eta = -\frac{1}{2\theta}, \quad \theta = -\frac{1}{2\eta}$$

and, consequently,

$$\underbrace{A(\eta)}_{B(\theta)=N\ln(\theta)} = N\ln\left(-\frac{1}{2\eta}\right) = -N\ln(-2\eta).$$

Therefore, the natural sufficient statistic $T(\boldsymbol{X}) = \sum_{n=0}^{N-1} X^2[n]$ has mean

$$\mathrm{E}_{\boldsymbol{X}|\eta}\left(\sum_{n=0}^{N-1} X^2[n] \,\middle|\, \eta\right) = \frac{\mathrm{d}A(\eta)}{\mathrm{d}\eta} = -\frac{N}{\eta} = 2N\theta$$

and variance

$$\begin{aligned}\mathrm{var}_{\boldsymbol{X}|\eta}\left(\sum_{n=0}^{N-1} X^2[n] \,\middle|\, \eta\right) &= \frac{\mathrm{d}^2 A(\eta)}{\mathrm{d}\eta^2} \\ &= \frac{N}{\eta^2} \\ &= 4N\theta^2.\end{aligned}$$

✻ COMMENTS. Examples of distributions in the exponential family: Gaussian with unknown mean or variance, exponential with unknown mean, Poisson with unknown mean, gamma, Bernoulli with unknown success probability, binomial with unknown success probability, multinomial with unknown cell probabilities.

Examples of distributions that are *not* members of the exponential family: Cauchy with unknown median, uniform with unknown support, F with unknown degrees of freedom.

References

Bickel, Peter J. and Kjell A. Doksum (2001). *Mathematical Statistics: Basic Ideas and Selected Topics*. 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Hero, Alfred O. (2015). *Statistical Methods for Signal Processing*. Lecture notes. Univ. Michigan, Ann Arbor, MI.