

# Quiz 11: define an exponential family and show that the parameters are asymptotically normal

Daria Fitisova

Ladislaus von Bortkiewicz Chair of Statistics  
C.A.S.E. – Center for Applied Statistics  
and Economics

Humboldt-Universität zu Berlin

<http://lwb.wiwi.hu-berlin.de>

<http://case.hu-berlin.de>



## What is exponential family?

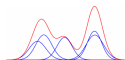
For a vector parameter  $\theta = (\theta_1, \dots, \theta_d)^T$ , a family of distributions is said to belong to the exponential family, if the probability density function (or PMF) can be written as

$$f(x | \theta) = h(x) \exp \left\{ \sum_{i=1}^s \eta_i(\theta) t_i(x) - a(\theta) \right\}$$

or more compact

$$f(x | \theta) = h(x) \exp \{ \eta(\theta)^T t(x) - a(\theta) \}.$$

Here,  $h(x)$  is the *base measure*,  $\eta(\theta)$  is called the *natural parameter*,  $t(x)$  is the *sufficient statistic* and  $a(\theta)$  is the *log - partition* or *log - normalizer*.



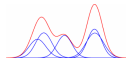
## Example: Poisson distribution

$$P(x | \theta) = h(x) \exp\{\eta(\theta)^T T(x) - a(\theta)\}$$

$$\begin{aligned} P(x | \lambda) &= \frac{\lambda^x e^{-\lambda}}{x!} = (x!)^{-1} \exp(-\lambda) \exp[x \log(\lambda)] \\ &= (x!)^{-1} \exp\{x \log \lambda - \lambda\}, \end{aligned}$$

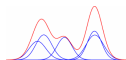
i.e. belonging to the exponential family with

- $\theta = \log \lambda$
- $h(x) = (x!)^{-1}$
- $t(x) = x$
- $a(\theta) = \lambda = e^\theta$



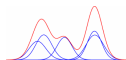
## Distributions in exponential family

- Gaussian:  $\mathbb{R}^p$
- Bernoulli: binary  $\{0,1\}$
- Binomial: counts of success/failure
- Multinomial: categorical
- Exponential:  $\mathbb{R}^+$
- Poisson:  $\mathbb{N}^+$
- Gamma:  $\mathbb{R}^+$
- ...



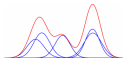
## Motivation

- provides a general framework for selecting a possible alternative parametrisation of the distribution
  - ▶ GLMs
  - ▶ Bayesian estimates in machine learning
- a number of important statistics can be derived at one stroke within the framework of exponential family



## Moments of exponential family

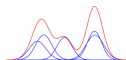
$$\begin{aligned}\frac{\partial a(\theta)}{\partial \theta} &= \frac{d}{d\theta} \left\{ \log \left( \int \exp\{\theta^T t(x)\} h(x) dx \right) \right\} \\ &= \frac{\frac{d}{d\theta} \int \exp\{\theta^T t(x)\} h(x) dx}{\int \exp\{\theta^T t(x)\} h(x) dx} \\ &= \frac{t(x) h(x) \exp\{\theta^T t(x)\} dx}{\int \exp\{\theta^T t(x)\} h(x) dx} \\ &= \frac{t(x) \exp\{\theta^T t(x)\} h(x) dx}{\exp\{-a(\theta)\}} \\ &= \int t(x) \exp\{\theta^T t(x) - a(\theta)\} h(x) dx \\ &= \mathbb{E}[t(x)]\end{aligned}$$



## Moments of exponential family

Likewise, it can be shown that:

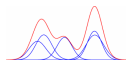
$$\frac{\partial^2 a(\theta)}{\partial \theta^2} = \text{Var}(t(x)) = \mathbb{E}[t(x)^2] - \mathbb{E}[t(x)]^2$$



## Example: moments of Poisson distribution

- ▣  $\theta = \log(\lambda)$ : canonical parameter
- ▣  $a(\theta) = \exp(\theta)$
- ▣  $\mathbb{E}[Y] = a'(\theta)$
- ▣  $\text{Var}[Y] = a''(\theta)$

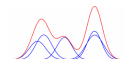
wherefrom follows that for Poisson distribution yields  $\mathbb{E}[Y] = \lambda$ ,  
 $\text{Var}[Y] = \lambda$





## Overview of some distributions in exponential family

Distrib.	$\theta$	$a(\theta)$	$\phi$	$E[Y]$	$V(\mu)$
$\text{Bin}(n, \pi)$	$\log(\frac{\pi}{1-\pi})$	$\log(1 + e^\theta)$	1	$\mu = n\pi$	$n\pi(1 - \pi) = \mu(1 - \mu/n)$
$\text{Po}(\mu)$	$\log(\mu)$	$\exp(\theta)$	1	$\mu$	$\mu$
$N(\mu, \sigma^2)$	$\mu$	$\frac{\theta^2}{2}$	$\sigma^2$	$\mu$	1
$\text{Gamma}(\mu, \nu)$	$-\frac{1}{\mu}$	$-\log(-\mu)$	$\frac{1}{\nu}$	$\mu$	$\mu^2$
$\text{IG}(\mu, \sigma^2)$	$-\frac{1}{2\mu^2}$	$-\sqrt{-2\theta}$	$\sigma^2$	$\mu$	$\mu^3$
$\text{NB}(\mu, \kappa)$	$\log(\frac{\kappa\mu}{1+\kappa\mu})$	$\frac{-1}{\kappa} \log(1 - \kappa e^\theta)$	1	$\mu$	$\mu(1 + \kappa\mu)$



## MLE of the mean parameter in exponential family distributions

$$\ell(\theta \mid D) = \log\left(\prod_{n=1}^N h(x_n)\right) + \theta^T \left(\sum_{n=1}^N t(x_n)\right) - Na(\theta).$$

Taking the gradient with respect to  $\theta$  gives:

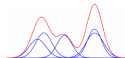
$$\nabla_{\theta} \ell = \sum_{n=1}^N t(x_n) - N \nabla_{\theta} a(\theta),$$

and setting to zero gives:

$$\nabla_{\theta} a(\hat{\theta}) = \frac{1}{N} \sum_{n=1}^N t(x_n).$$

Finally, defining  $\mu := \mathbb{E}[t(x)]$ , we obtain:

$$\hat{\mu}_{ML} = \frac{1}{N} \sum_{n=1}^N t(x_n)$$

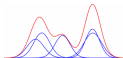


## Recap: asymptotic normality

Assume a sequence of RVs  $\{X_n\}$  with mean  $\mu_n$  and variance  $\sigma_n^2$ .  $\{X_n\}$  is asymptotically normal if

$$\frac{X_n - \mu_n}{\sigma_n} \xrightarrow{D} N(0, 1)$$

Intuition: as we get more and more data, averages of random variables behave like normally distributed random variables



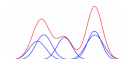
## AN of MLE in exponential family: Poisson distribution example

$$\hat{\mu}_{ML} = \frac{1}{N} \sum_{n=1}^N t(X_n)$$

By the assumption  $X_i$  are i.i.d., and since  $t(X_i) = X_i$  (consequently,  $\hat{\mu}_{ML} = \bar{X}_n$ ) we can apply the iid version of the Central Limit Theorem directly to  $\hat{\mu}_{ML}$ .

Since  $\mathbb{E}(X_i) = \text{Var}(X_i) = \theta < \infty$  for all  $\theta$  in the parameter space, it follows that

$$\sqrt{n}(\hat{\mu}_{ML} - \theta) / \sqrt{\theta} \xrightarrow{D} N(0, 1)$$



## AN of MLE in exponential family: general case

The proof for the general case is quite complicated. Here is the guideline:

- Continuous mapping theorem
- Central limit theorem
- Delta theorem

For more details one can see:

<http://www.stat.purdue.edu/~dasgupta/mle.pdf>

