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://lvb.wiwi.hu-berlin.de
http://case.hu-berlin.de

What is exponential family?

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

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

or more compact

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

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

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

$$P(x \mid \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} = (x!)^{-1} exp(-\lambda) exp[xlog(\lambda)]$$
$$= (x!)^{-1} exp\{xlog\lambda - \lambda\},$$

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:

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

Example: moments of Poisson distribution

- θ=log(λ): canonical parameter
- a(θ)=exp(θ)
- $\mathbb{E}$[Y]=a'(θ)
- Var[Y]=a''(θ)

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

Overview of some distributions in exponential family

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

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 θ 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 μ:=$\mathbb{E}$[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 μ_n and variance σ_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 $\mu_{\hat{ML}}$.
Since $\mathbb{E}(X_i)=Var(X_i)=\theta < \infty$ for all θ 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