National Technical University of Ukraine
“Igor Sikorsky Kyiv Polytechnic Institute”
Institute of Physics and Technology

Lecture 7
Generative Models for Discrete Data. Generalized Linear Models

Dmytro Progonov,
PhD, Associate Professor,
Department Of Physics and Information Security Systems

Content

- The exponential family;
- MLE for the exponential family;
- Bayes for the exponential family;
- Generalized Linear Models (GLM);
- Generalized Linear Mixed Models (GLMM).

The exponential family

Before defining **the exponential family**, we mention several reasons why it **is important**:

- It can be shown that, under certain regularity conditions, the exponential family is the only **family of distributions with finite-sized sufficient statistics**;
- The exponential family is only **family of distributions for which conjugate priors exist**, which simplifies the computation of the posterior;
- The exponential family can be shown to be **the family of distributions that makes the least set of assumptions** subject to some user-chosen constrains;
- The exponential family is at the **core of generalized linear models**;
- The exponential family is at the **core of variational inference**.

Definition of exponential family

A pdf or pmf $p(\mathbf{x}|\boldsymbol{\theta})$ for $\mathbf{x} \in \mathcal{X}^m$ and $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$ is said to be in the exponential family if it is on the form:

$$\boxed{p(\mathbf{x}|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \exp[\boldsymbol{\theta}^T \phi(\mathbf{x})] = h(\mathbf{x}) \exp[\boldsymbol{\theta}^T \phi(\mathbf{x}) - A(\boldsymbol{\theta})],}$$

where

$\boldsymbol{\theta}$ – natural (canonical) parameters;

$\phi(\mathbf{x}) \in \mathbb{R}^d$ – vector of sufficient statistics;

$Z(\boldsymbol{\theta}) = \int_{\mathcal{X}^m} h(\mathbf{x}) \exp[\boldsymbol{\theta}^T \phi(\mathbf{x})]\, d\mathbf{x}$ – partition function;

$A(\boldsymbol{\theta}) = \log[Z(\boldsymbol{\theta})]$ – log partition (cumulant) function;

$h(\mathbf{x})$ – scaling constant;

$\boldsymbol{\eta}(\boldsymbol{\theta})$ – function that maps parameters $\boldsymbol{\theta}$ to the canonical parameters.

Example of exponential family

The Bernoulli for $x \in \{0; 1\}$ can be written in exponential family form as follows:

$$\boxed{Ber(x|\mu) = (1-\mu)\exp\left[x\log\left[\frac{\mu}{1-\mu}\right]\right],}$$

where

$\mu = \mathrm{sigm}[\theta] = \frac{1}{1+e^{-\theta}}$ – mean parameter;

$\phi(x) = x$ – vector of sufficient statistics;

$\theta = \log\left[\frac{\mu}{1-\mu}\right]$ – natural (canonical) parameters;

$Z(\theta) = \frac{1}{1-\mu}$ – partition function.

MLE for the exponential family

The likelihood of an exponential family model has the form:

$$\boxed{p(\mathcal{D}|\boldsymbol{\theta}) = \left[\prod_{i=1}^{N} h(\mathbf{x}_i)\right] g(\boldsymbol{\theta})^N \exp\left[\boldsymbol{\eta}(\boldsymbol{\theta})^T \sum_{i=1}^{N} \phi(\mathbf{x}_i)\right],}$$

where

$$\phi(\mathcal{D}) = \left[\sum_{i=1}^{N} \phi_1(\mathbf{x}_i), \sum_{i=1}^{N} \phi_2(\mathbf{x}_i), \cdots, \sum_{i=1}^{N} \phi_K(\mathbf{x}_i), \right].$$

The ***Pitman-Koopman-Darmois theorem*** states that, under certain regularity conditions, the exponential family is the only family of distributions with finite sufficient statistics.

Bayes for the exponential family

Exact Bayesian statistics is considerably simplified if the prior is conjugate to the likelihood. This means that the prior $p(\boldsymbol{\theta}|\boldsymbol{\tau})$ has the same form as likelihood $p(\mathcal{D}|\boldsymbol{\theta})$. For this to make sense, we require that the likelihood have finite sufficient statistics, so that we can write $p(\mathcal{D}|\boldsymbol{\theta}) = p(\mathbf{s}(\mathcal{D})|\boldsymbol{\theta})$.

Likelihood of the exponential family is given by

$$\boxed{p(\mathcal{D}|\boldsymbol{\theta}) \propto g(\boldsymbol{\theta})^N \exp[\boldsymbol{\eta}(\boldsymbol{\theta})^T \mathbf{s}_N]}$$

Prior – $\boxed{p(\boldsymbol{\theta}|v_0, \boldsymbol{\tau}_0) \propto g(\boldsymbol{\theta})^{v_0} \exp[\boldsymbol{\eta}(\boldsymbol{\theta})^T \boldsymbol{\tau}_0]}$

Posterior – $\boxed{p(\boldsymbol{\theta}|\mathcal{D}) = p(\boldsymbol{\theta}|v_0 + N, \boldsymbol{\tau}_0 + \mathbf{s}_N)}$

where $\mathbf{s}_N = \sum_{i=1}^{N} \mathbf{s}(\mathbf{x}_i)$; v_0 – size of prior pseudo-data; $\overline{\boldsymbol{\tau}_0}$ – mean of the sufficient statistics on this pseudo-data; $\boldsymbol{\tau}_0 = v_0 \overline{\boldsymbol{\tau}_0}$.

Generalized Linear Models (GLM)

These are models in which *the output density is in the exponential family*, and in which *the mean parameters are a linear combination of the inputs*, passed through a possibly non-linear function.

Distribution	Link function $g(\mu)$	Convert function from mean to natural parameters $\theta = \psi(\mu)$	Mean parameter $\mu = \psi^{-1}(\theta) = \mathbb{E}[y]$
$\mathcal{N}(\mu, \sigma^2)$	identity	$\theta = \mu$	μ=θ
$Bin(N, \mu)$	logit	$\theta = \log\left[\frac{\mu}{1-\mu}\right]$	$\mu = \mathrm{sigm}(\theta)$
$Poi(\theta)$	log	$\theta = \log[\mu]$	$\mu = e^{\theta}$

Generalized Linear Mixed Models (GLMM)

Suppose we generalize the multi-task learning scenario to allow the response to include information at the group level $(\mathbf{x}_j)$ as well as at the item level $(\mathbf{x}_{ij})$. Similarly, we can allow the parameters to vary across group$(\boldsymbol{\beta}_j)$ or to be tied across group $\boldsymbol{\alpha}$. This gives rise to the following model:

$$\boxed{\mathbb{E}[y_{ij}|\mathbf{x}_{ij}, \mathbf{x}_j] = g\left(\phi_1(\mathbf{x}_{ij})^T\boldsymbol{\beta}_j + \phi_2(\mathbf{x}_j)^T\acute{\boldsymbol{\beta}}_j + \phi_3(\mathbf{x}_{ij})^T\boldsymbol{\alpha} + \phi_4(\mathbf{x}_j)^T\acute{\boldsymbol{\alpha}}\right)}$$

Frequentist call the term $\boldsymbol{\beta}_j$ **random effects**, since they vary randomly across groups, but they call $\boldsymbol{\alpha}$ a **fixed effect**, since it is viewed as a fixed but unknown constant.

A model with both fixed and random effects is called a **mixed model**. If $p(y|\mathbf{x})$ is GLM, the overall model is called a GLMM.

Conclusion

- The definition of exponential family was presented;
- Procedures for estimation the parameters of the exponential family distribution were shown;
- Generalized Linear and Mixed Models were considered.