STAT-510 **2017 Fall**

Lecture 2: Families of Distributions

Instructor: Xiaohui Chen Scribe: Yutong Dai Last Modified: 2018-02-16

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

2.1 Location-Scale Family

Definition 2.1 Let $f(x)$ be any pdf. Then for any $-\infty < \mu < +\infty$ and any $\sigma > 0$, the family of pdfs $\{\frac{1}{\sigma}f(\frac{x-\mu}{\sigma})| -\infty < \mu < +\infty, \sigma > 0\}$ is called the location-scale familiy with standard pdf $f(x)$; μ is called the location parameter and σ is called the scale parameter.

Remark:

1. $\int \frac{1}{\sigma}f(\frac{x-\mu}{\sigma})dx = 1$, therefore, $\frac{1}{\sigma}f(\frac{x-\mu}{\sigma})$ are distributions.
2. $\sigma > 1$ means to strech the pdf while $\sigma < 1$ means to contract the pdf.
3. The point of defining the location-scale family is to state that we can generate a family of pdfs with any pdf $f(x)$ by introducing scale or location parameters.

Theorem 2.2 Let $f(x)$ be any pdf, $\mu \in R$ and $\sigma > 0$. Then Z is a random variable with pdf $\frac{1}{\sigma}f(\frac{z-\mu}{\sigma})$ $\iff$ $Z = \sigma X + \mu$.

2.2 Exponential Families

Definition 2.3 A family of pdfs or pmfs is called an exponential family if it can be expressed as

$$f(x;\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})exp(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)),$$

where $h(x) > 0$ and $\forall i, t_i(x)$ is real-valued function and doesn't depend on $\boldsymbol{\theta} \in R^d$, $d \leq k$. Also, $c(\boldsymbol{\theta}) \geq 0$ and $\forall i, w_i(\boldsymbol{\theta})$ is real-valued function and doesn't depend on x.

Claim 2.4 If $f(x;\theta)$ is in the exponential family, then the support of $f(x;\theta)$, $\{x \in \mathcal{X}; f(x;\theta) > 0\}$, doesn't depend on parameter θ. It only depends on the set $\{x|h(x) > 0\}$.

Remark

1. Noraml, Gamma, Beta, Binomial, Possion, negative Binominal distributions are all in exponential family.
2. Some distributions in location family are not in the exponential family, for example, $f(x;\theta) = \frac{1}{\theta}exp(1 - \frac{x}{\theta})I(x > \theta), \theta > 0$.

An exponential family is sometimes reparameterized as

$$f(x;\eta) = c^*(\eta)h(x)exp(\sum_{i=1}^{k} \eta_i t_i(x)),$$

where $h(.), t_i(.)$ are the same as in the original parameterization. And $c^*(\eta)$ is the parameter that ensures $\int f(x;\eta)dx = 1$. The set

$$H = \{(\eta_1, ..., \eta_k)| \int_{-\infty}^{+\infty} h(x)exp(\sum_{i=1}^{k} \eta_i t_i(x))dx < +\infty\},$$

is called the *natural parameter space* for the family. Since the original $f(x;\boldsymbol{\theta})$ is a pdf/pmf in exponential family, $(\eta = (w_1(\boldsymbol{\theta}), w_2(\boldsymbol{\theta}), ..., w_k(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta)$ must be a subset of the natural parameter space.

Definition 2.5 A curved exponential family is a family of densities of the form $f(x;\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})exp(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x))$ for which the dimension of the vector $\boldsymbol{\theta}$ is equal to $d \leq k$. If $d = k$, then the family is a **full exponential family**. If $d < k$, it is **curved exponential family**.