STAT 345/445 Lecture 12

Groups of Families of Distributions – Sections 3.4 and 3.5

Chebychev's Inequality – Section 3.6

Note: We will skip the rest of Section 3.6, for now.

1. Exponential Families
2. Location - scale families
3. Chebychev's Inequality

Groups of families

- Have seen many families of distributions
  - Family of Normal distributions, Family of Poisson distribution etc.
- We will now define two groups of families
  - **Exponential families**
  - **Location-scale families**
- Use: prove properties for all families of distributions in a group
  - Will see more of that in STAT 346/446
- Example: Theory for Generalized linear models (GLMs) is derived for all exponential families
  - Logistic regression, Poisson regression, etc.

Exponential Families

Definition

A family of pdfs or pmfs indexed by parameter(s) $\boldsymbol{\theta}$ is called an **exponential family** if it can be written as

$$f(x \mid \boldsymbol{\theta}) = h(x)\, c(\boldsymbol{\theta})\, \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(x)\right) \qquad \forall x \in \mathbb{R}$$

where

- $h(x), t_1(x), \ldots, t_k(x)$ are functions of x only (not $\boldsymbol{\theta}$)
- $c(\boldsymbol{\theta}), w_1(\boldsymbol{\theta}), \ldots, w_k(\boldsymbol{\theta})$ are functions of $\boldsymbol{\theta}$ only (not x)
- $h(x) \geq 0\ \forall x$ and $c(\boldsymbol{\theta}) \geq 0\ \forall \boldsymbol{\theta}$

Examples of exponential families

- $\mathrm{N}(\mu, \sigma^2)$
- $\mathrm{Binomial}(n, p)$ if n is known (fixed)
- $\mathrm{Expo}(\beta)$

done on whiteboard

Indicator function: A handy tool to get more compact expressions of pdf/pmf:

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

- Example of a family that is *not* and exponential family: $\mathrm{Uniform}(a, b)$

done on whiteboard

Mean and variance for exponential families

Theorem

If X is a random variable with a pdf or pmf from an exponential family then

$$\mathrm{E}\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(X)\right) = -\frac{\partial}{\partial \theta_j} \log\left(c(\boldsymbol{\theta})\right)$$

$$\mathrm{Var}\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(X)\right) = -\frac{\partial^2}{\partial \theta_j^2} \log\left(c(\boldsymbol{\theta})\right) - \mathrm{E}\left(\sum_{i=1}^{k} \frac{\partial^2 w_i(\boldsymbol{\theta})}{\partial \theta_j^2} t_i(X)\right)$$

- Example: $\mathrm{Expo}(\beta)$ done on whiteboard

Curved vs. full exponential families

A pdf/pmf from an exponential family:

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

- Often the dimension of $\boldsymbol{\theta}$ is equal to k - but not always

Definition: Curved or Full Expo Families

If we can write $f(x)$ such that $k = d$ where d is the dimension of the vector $\boldsymbol{\theta}$, the familiy is called a **full exponential family**. A **curved exponential family** is an exponential family for which $d < k$.

- Example: $\mathrm{N}(\theta, \theta^2)$
- Some properties (see e.g. chapter 6) can only be shown for *full* exponential families

Location-scale families

- First, a handy theorem about shifting and re-scaling pdfs:

Theorem

Let $f(x)$ be a pdf and let $\mu \in \mathbb{R}$, $\sigma > 0$ be constants. Then

$$g(x \mid \mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

is also a pdf.

proof... done on whiteboard

Location-scale families

Definition

Let $f(x)$ be a pdf (sometimes called the *standard pdf*)

(i) Set $g(x \mid \mu) = f(x - \mu)$. Then $\{g(x \mid \mu) : \mu \in \mathbb{R}\}$ is called a **location family**

(ii) Set $g(x \mid \sigma) = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$. Then $\{g(x \mid \sigma) : \sigma > 0\}$ is called a **scale family**

(iii) Set $g(x \mid \mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$. Then $\{g(x \mid \mu, \sigma) : \mu \in \mathbb{R}, \sigma > 0\}$ is called a **location-scale family**

μ is called a **location parameter** and σ is called a **scale parameter**

- Example: $\mathrm{N}(\mu, \sigma^2)$ done on whiteboard

Location-scale families

- If support of $f(x)$ is not $\mathbb{R}$ then the support of $g(x \mid \mu, \sigma)$ will depend on μ and σ
- Example: Define a location-scale family with $f(x)$ the pdf for $\mathrm{Uniform}(a, b)$

done on whiteboard

Location-scale families

One use of location-scale families:

- Probabilities for any location-scale pdf can be calculated by transforming to the standard pdf

Theorem

Let $g(\cdot \mid \mu, \sigma)$ be a pdf from a location-scale family with standard pdf $f(\cdot)$.

(a) If $X \sim g(x \mid \mu, \sigma)$ then $Z = \frac{X-\mu}{\sigma} \sim f(z)$

(b) If $Z \sim f(z)$ then $X = \sigma Z + \mu \sim g(x \mid \mu, \sigma)$

- Examples: Normal distribution, Uniform distribution ...

Chebychev's Inequality

Theorem: Chebychev's Inequality

Let X be a random variable and let $g(x)$ be a non-negative function. Then for any $k > 0$

$$P(g(X) \geq k) \leq \frac{E(g(X))}{k}$$

proof... done on whiteboard

Example of Chebychev's Inequality

- Let X be a random variable with mean $\mu = E(X)$ and variance $\sigma^2 = V(X)$. Consider

$$g(x) = \frac{(x-\mu)^2}{\sigma^2}$$

what does Chebychev's inequality imply?

$$P(|X-\mu| \geq t\sigma) \leq \frac{1}{t^2}$$

or equiv: $P(|X-\mu| < t\sigma) \leq 1-\frac{1}{t^2}$

$t=2$: $P(|X-\mu| < 2\sigma) \leq 1-\frac{1}{2^2} = 0.75$

$t=6$: $P(|X-\mu| < 6\sigma) \leq 1-\frac{1}{36} \approx 0.9722$