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 $c(\boldsymbol{\theta}) \geq 0 \; \forall \boldsymbol{\theta}$

$$f(x) = \frac{1}{\sqrt{2\pi}\,\Delta} e^{-\frac{1}{2\Delta^2}(x^2 - 2x\mu + \mu^2)}$$

$$= \frac{1}{\sqrt{2\pi}\,\Delta} e^{-\frac{1}{2\Delta^2}x^2 + \frac{x\mu}{\Delta^2}} e^{-\frac{\mu^2}{2\pi^2}}$$

$$= \frac{1}{\sqrt{2\pi}\,\Delta} e^{-\mu^2/2\Delta^2} \cdot e^{-\frac{1}{2\Delta^2}x^2 + x\frac{\mu}{\Delta^2}}$$

$$= \underbrace{\frac{1}{\sqrt{2\pi}}}_{h(x)} \underbrace{\frac{1}{\Delta} e^{-\mu^2/2\Delta^2}}_{= C(\mu,\Delta)} e^{\overbrace{-\frac{1}{2\Delta^2}}^{w_1(\mu,\Delta)}\overbrace{x^2}^{t_1(x)} + \overbrace{x}^{t_2(x)}\overbrace{\frac{\mu}{\Delta^2}}^{w_2(\mu,\Delta)}}$$

$(k = 2)$

$$h(x) = I_{\mathbb{R}}(x)\frac{1}{\sqrt{2\pi}}$$

Examples of exponential families

- $\mathrm{N}(\mu, \sigma^2)$ is an exponential families
- $\mathrm{Binomial}(n, p)$ if n is known (fixed)
- $\mathrm{Expo}(\beta)$ is an exponential families, pdf: $f(x) = \frac{1}{\beta} e^{-x/\beta}$, $I_{[0,\infty)}(x)$

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

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

$I_{\mathbb{R}}(x)$

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

$X \sim \text{Binomial}(n, p)$ n is fixed (i.e. not a parameter)

Define $A = \{0, 1, 2, \ldots n\}$

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} I_A(x) = \binom{n}{x} I_A(x) e^{\ln(p^x (1-p)^{n-x})}$$

$$= \binom{n}{x} I_A(x) \exp\left(\ln(p^x) + \ln((1-p)^{n-x})\right) \ldots (*)$$

Recall: $\ln(ab) = \ln(a) + \ln(b)$ and $\ln(a^b) = b\ln(a)$

$$(A) = \binom{n}{x} I_A(x) \exp\left(x\ln(p) + (n-x)\ln(1-p)\right)$$

$$= \binom{n}{x} I_A(x) \exp\left(x(\ln p - \ln(1-p)) + n\ln(1-p)\right)$$

$$= \binom{n}{x} I_A(x)(1-p)^n \exp\left(x\ln\frac{p}{1-p}\right)$$

Show that $Expo(\beta)$ is an exponential family

pdf: $f(x) = \frac{1}{\beta} e^{-x/\beta} I_{(0,\infty)}(x)$

set $h(x) = I_{(0,\infty)}(x)$

$C(\beta) = \frac{1}{\beta}$

$W_i(\beta) = -\frac{1}{\beta}$

$t_i(x) = x$

then

$f(x) = h(x) C(\beta) \exp(W_i(\beta) t(x))$

$k = 1$

Uniform(a, b) is not an exponential family:

$f(x) = \frac{1}{b-a} \underbrace{I_{[a,b]}(x)}_{\text{function of both } x \text{ and } a, b}$ pars: a and β

can't be written as

either $c(a,b)h(x)$ or $\exp(w(a,b)t(x))$

In general: If the support of the distribution depend on a parameter it is not an exp family

Eg: if both n and p are unknown in

Binomial(n, p), it is no longer an exp family

(saw last time)

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: $\mathrm{Expo}(\beta)$

Mean and variance for Expo(β) $\log = \ln$

$$E\left(\frac{\partial w_1(\beta)}{\partial \beta} t_1(x)\right) = -\frac{\partial}{\partial \beta}\log(c(\beta)) = -\frac{\partial}{\partial \beta}\log\frac{1}{\beta}$$

$$= \frac{\partial}{\partial \beta}\log\beta$$

$$E\left(\frac{\partial}{\partial \beta}\left(-\frac{1}{\beta}\right)X\right) = E\left(\frac{1}{\beta^2}X\right) = \frac{1}{\beta^2}E(X)$$

$$= \frac{1}{\beta}$$

$$\frac{1}{\beta^2}E(X) = \frac{1}{\beta} \qquad E(X) = \beta$$

About Var:

$$-\frac{\partial^2}{\partial\beta^2} C(\beta) = \frac{\partial}{\partial\beta}\frac{1}{\beta} = -\frac{1}{\beta^2} \quad \text{and}$$

$$E\left(\frac{\partial^2}{\partial\beta^2} W_1(\beta)\, t_1(X)\right) = E\left(\frac{\partial}{\partial\beta}\frac{1}{\beta^2}X\right) = E\left(-\frac{2}{\beta^3}X\right) = -\frac{2}{\beta^3}E(X)$$

$\Rightarrow$ by (ii): $Var\left(\frac{1}{\beta^2}X\right) = -\frac{1}{\beta^2} + \frac{2}{\beta^2}\beta$

$$Var(X) = \frac{\beta^4}{\beta^2} = \beta^2$$

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... $f(x)$ is a pdf. Show that

$g(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, $\mu \in \mathbb{R}$, $\sigma > 0$, is a pdf.

$g(x) \geq 0 \ \forall x$ since $f(x) \geq 0 \ \forall x$ and $\delta > 0$

$$\int_{-\infty}^{\infty} g(x)\, dx = \int_{-\infty}^{\infty} \frac{1}{\delta} f\left(\frac{x-\mu}{\delta}\right) dx = \int_{-\infty}^{\infty} \frac{1}{\delta} f(u) \delta\, du$$

Substitution $u = \dfrac{x-\mu}{\delta}$

$$= \int_{-\infty}^{\infty} f(u)\, du = 1$$

$$\frac{du}{dx} = \frac{1}{\delta} \Rightarrow dx = \delta\, du \quad \text{and} \quad x = \mu + u\delta$$

Rang: $-\infty < x < \infty$

$-\infty < \dfrac{x-\mu}{\delta} < \infty$

$\Rightarrow g(x)$ is a pdf

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: $\mathrm{N}(\mu, \sigma^2)$ is a location-scale family with the standard pdf $N(0,1)$: $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} I_{(-\infty,\infty)}(x)$

$$g(x) = \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2/2} I_{(-\infty,\infty)}\left(\frac{x-\mu}{\sigma}\right)$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2}} I_{(-\infty,\infty)}(x)$$

$$= \text{pdf of } N(\mu, \sigma^2)$$

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 $\mathrm{Uniform}(a, b)$

Uniform(a, b), a and b fixed is the standard pdf

$$f(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x)$$

$$g(x) = \frac{1}{\sigma} \frac{1}{b-a} \mathbb{1}_{[a,b]}\left(\frac{x-\mu}{\sigma}\right)$$

$$= \frac{1}{\sigma(b-a)} \mathbb{1}_{[\sigma a+\mu, \sigma b+\mu]}(x), \Rightarrow \text{pdf of uniform}[\sigma a+\mu, \sigma b+\mu)$$

Note:

$$a \le \frac{x-\mu}{\sigma} \le b$$

$$\Rightarrow \sigma a+\mu \le x \le \sigma b+\mu$$

Notice: μ is not the mean of $g(x)$.

$$\frac{\Delta a+\mu+\Delta b+\mu}{2} = \mu + \frac{\Delta a+\Delta b}{2} = \mu + \Delta \underbrace{\frac{a+b}{2}}_{\text{mean of } f(x)}$$

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... First Note:

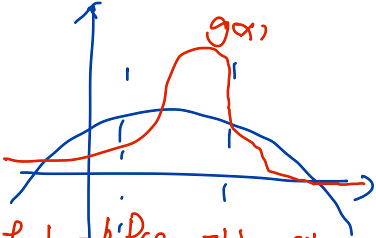
$$E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

$$P(x \geq k) = \int_{k}^{\infty} f(x)\, dx$$

or $P(X \in A) = \int_A f(x)\, dx$

$$P(g(x) \geq k) = P(X \in \{x \in \mathbb{R} : g(x) \geq k\})$$

$$= \int_{\{x : g(x) \geq k\}} f(x)\,dx$$



$$E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x)\,dx \geq \int_{\{x : g(x) \geq k\}} g(x) f(x)\,dx$$

$$\geq \int_{\{x : g(x) \geq k\}} k f(x)\,dx = k\,P(g(x) \geq k) \cdots (*)$$

integrating over a smaller range.

$\{x : g(x) \geq k\}$

$g(x) f(x) \geq k f(x)$ for all $x \in \{x : g(x) \geq k\}$

$(*)$

$\Downarrow$

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

Example of Chebychev's Inequality

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

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

is a non-negative function.

what does Chebychev's inequality imply?

$$\underbrace{P\left(\frac{(x-\mu)^2}{\sigma^2} \geq k\right)}_{\parallel} \leq \frac{1}{k} E\left(\frac{(x-\mu)^2}{\sigma^2}\right)$$

$$P((x-\mu)^2 \geq k\sigma^2) \qquad \frac{1}{k\sigma^2} E((x-\mu)^2) = \frac{1}{k\sigma^2}\sigma^2 = \frac{1}{k} \quad \cdots (*)$$

$$= P(|X-\mu| \geq \sqrt{k}\,\sigma)$$

prob that X is more than $\sqrt{k}$ stand dev from its mean

(*) This is usually presented with $t = \sqrt{k}$

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

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

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

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

For $X \sim N(\mu, \sigma)$ we get

$$P(|x-\mu| < 2\delta) \simeq 0.95$$