Math 502AB - Lecture 9

Dr. Jamshidian

September 20, 2017

1 Lecture - Part 1

1.1 Exponential Family, Continued

Recall:

The exponential family of distribution is of the form:

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

1.1.1 The Natural Paramaterization

A reparameterization of the exponential family where the density is written in the form:

$$f(x|\eta) = h(x)c^*(\eta)\exp\left[\sum_{i=1}^{k} \eta_i t_i(x)\right]$$

Where $h(x)$ and $t_i(x)$ are the same as before, but the parameters enter linearly in the sum. The difference is in the summation part. In the *natural parameterization*, we have parameters expressed in linear functions (ie. α instead of $\log\alpha$).

The set H as defined below is referred to as the *natural parameter space*:

$$H = \left\{(\eta_1, ..., \eta_p) : \int_{-\infty}^{\infty} h(x)\exp\left[\sum_{i=1}^{k} \eta_i t_i(x)\right] dx < \infty\right\}$$

This means that we have:

$$c^*(\eta) = \left[\int_{-\infty}^{\infty} h(x)\exp\left\{\sum_{i=1}^{k} \eta_i t_i(x)\right\} dx\right]^{-1}$$

Examples

1. $X \sim gamma(\alpha, \beta)$

$$f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \frac{1}{x} \exp\left\{\alpha \log x - \frac{1}{\beta} x\right\}$$

Since $\frac{1}{\beta}$ appears in the summation portion, this is *not* a natural parameterization. We then rewrite the expression to find the natural parameterization:

$$\eta_1 = \alpha \quad \text{and} \quad \eta_2 = \frac{1}{\beta}$$

$$\Rightarrow f_X(x) = \frac{\eta_2^{\eta_1}}{\Gamma(\eta_1)} \cdot \frac{1}{x} \cdot \exp\{\eta_1 \log x - \eta_2 x\}$$

Thus, our natural parameterization is:

$$H = \left\{\left(\alpha, \frac{1}{\beta}\right) : \quad \alpha > 0, \beta > 0\right\}$$

2. Use the above parameterization to obtain $E(X)$ and $E[\log X]$ for $x \sim gamma(\alpha, \beta)$

Recall:

$$E\left[\sum_{i=1}^{k} \frac{\partial w_i(\eta)}{\partial \eta_j} t_i(x)\right] = -\frac{\partial}{\partial \eta_j} \log c(\eta)$$

In general, expected value might be very difficult to compute using integration. This is the motivation behind taking the derivative.

$$E\left[\frac{\partial w_1(\eta)}{\partial \eta_1} t_1(x) + \frac{\partial w_2(\eta)}{\partial \eta_2} t_2(x)\right] = E[\log x]$$

$$\begin{aligned}
LHS &= E\left[\frac{\partial w_1(\eta)}{\partial \eta_1} t_1(x) + \frac{\partial w_2(\eta)}{\partial \eta_2} t_2(x)\right] = E[\log x] \\
RHS &= -\frac{\partial}{\partial \eta_1}\left[\log\left(\frac{\eta_2^{\eta_1}}{\Gamma(\eta_1)}\right)\right] \\
&= -\frac{\partial}{\partial \eta_1}\left[\eta_1 \log \eta_2 - \log \Gamma(\eta_1)\right] \\
&= -\log \eta_2 + \frac{\Gamma'(\eta_1)}{\Gamma(\eta_1)}
\end{aligned}$$

$$\begin{aligned}
LHS &= -E[X] \\
RHS &= -\frac{\partial}{\partial \eta_2}\left[\eta_1 \log \eta_2 - \log \Gamma(\eta_1)\right] \\
&= -\frac{\eta_1}{\eta_2} = -\alpha\beta
\end{aligned}$$

3. $X \sim N(\mu, \sigma^2)$

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x^2 - 2\mu x + \mu^2)\right\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\mu^2}{2\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x\right\} \end{aligned}$$

We note that:

$$\eta_1 = \frac{1}{\sigma^2} \quad \eta_2 = \frac{\mu}{\sigma^2} \Rightarrow \sigma^2 = \frac{1}{\eta_1} \quad \mu = \frac{\eta_2}{\eta_1}$$

We can now write that

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \eta_1^{0.5} e^{0.5\frac{\eta_2^2}{\eta_1}} \exp\left\{-\frac{1}{2}\eta_1 x^2 + \eta_2 x\right\}$$

$$\begin{aligned} LHS &= E\left[\sum_{i=1}^{k} \frac{\partial w_i(\eta)}{\partial \eta_j} t_i(x)\right] \\ &= E[X], \text{ since it is the natural parameter here} \\ RHS &= -\frac{\partial}{\partial \eta_j} \log c(\eta) \\ &= -\frac{\partial}{\partial \eta_2}\left[\frac{1}{2}\log \eta_1 - \frac{1}{2}\frac{\eta_2^2}{\eta_1}\right] = \frac{\eta_2}{\eta_1} = \mu \end{aligned}$$

1.2 Curved Exponential Families

1.2.1 Full Exponential vs Curved Exponential Families

A *curved exponential* family of distributions is a family with densities of the form

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

for which $\dim \theta < k$. If $\dim(\theta) = k$ then the family is called *full exponential family.*

Examples:

(a) $X \sim \gamma(\alpha, \beta)$ which has $k = 2$. Since $\theta = \{(\alpha, \beta) : \quad \alpha > 0, \beta > 0\}$ this is a *full exponential family.*

Consider a special case where $\alpha = \beta$. In this case, $k = 2$ still, but $\dim \theta = 1$.

(b) $X \sim \chi^2_{(p)} \sim gamma(\frac{p}{2}, 2)$

2 Lecture - Part 2

Quick Note: There has been one part of $h(x)$ that has been omitted. Say for example, we are writing the *pdf* for a distribution, let's say for the *gamma*:

$$f(x|\theta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \frac{1}{x} \exp\left\{\alpha \log x - \frac{1}{\beta} x\right\}$$

Now, this is supposed to be $\forall x$, but this is *non-zero* only when $x > 0$. So we add an indicator function:

$$f(x|\theta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \frac{1}{x} \exp\left\{\alpha \log x - \frac{1}{\beta} x\right\} \mathcal{I}_{[X \geq 0]}$$

Where $\mathcal{I}$ is the indicator function, and gets absorbed by $h(x)$.

$$h(x) = \frac{1}{x} \mathcal{I}_{[X \geq 0]}$$

2.1 Location Scale Families (Section 3.5)

The idea is to start with a standard form of a distribution and add parameters to *rescale* and *shift* the density or pmf. We have seen this before in the *standard normal* distribution $Z \sim N(0, 1)$. By adding μ and σ^2, you change the location and scale.

Theorem:

Let $f(x)$ be any *pdf.* Then for any μ and $\sigma > 0$:

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

To show this, we need to show that $g(x) > 0, \quad \forall x$ and that the integral over the domain is equal to 1.

Proof:

$g(x|\mu, \sigma) \geq 0 \quad \forall x$ is obvious.

$$\int_{-\infty}^{\infty} g(x|\mu,\sigma)dx = \int_{-\infty}^{\infty} \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) dx$$
$$\text{Letting } u = \frac{x-\mu}{\sigma} \Rightarrow du = \frac{1}{\sigma}dx$$
$$\Rightarrow \int_{-\infty}^{\infty} f(u)du = 1$$

Definitions:

1. **Location Family:** Let $f(x)$ be a *pdf*. Then the family of *pdf*'s $f(x-\mu)$, indexed by the parameter μ is called the **location family** with standard *pdf* $f(x)$. μ is referred to as the location parameter.

2. **Scale Family:** The family $\frac{1}{\sigma}f(\frac{x}{\sigma})$ is called a **scale family** for the *pdf* $f(x)$, with scale σ

3. **Location-Scale Family** The family $\frac{1}{\sigma}f(\frac{x-\mu}{\sigma})$ is called a **location-scale family** for the *pdf* $f(x)$ with scale σ and location parameter μ

Theorem:

Let f_Z be any *pdf*, and let $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$. Then x has density $f_X(x) = \frac{1}{\sigma}f_Z(\frac{x-\mu}{\sigma})$ **iff** $x = \mu + \sigma Z$, where Z has density f_Z

Proof

($\Leftarrow$): Suppose $x = \mu + \sigma Z$. Then:

$$\begin{aligned} F_X(x) = P(X \le x) &= P(\mu + \sigma Z \le x) \\ &= P\left(Z \le \frac{x-\mu}{\sigma}\right) \\ &= F_Z\left(\frac{x-\mu}{\sigma}\right) \\ f_X(x) &= \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

($\Rightarrow$): Suppose that

$$f_X(x) = \frac{1}{\sigma}f_Z(\frac{x-\mu}{\sigma})$$

We need to show:

$$P(X \le a) = P(\mu + \sigma Z \le a), \quad \forall a$$

$$
\begin{aligned}
P(X \le a) &= \int_{-\infty}^{a} f_X(x)dx \\
&= \int_{-\infty}^{a} \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right) dx \\
&\text{Let } z = \frac{x-\mu}{\sigma} \Rightarrow dz = \frac{1}{\sigma}dx \\
&= \int_{-\infty}^{\frac{a-\mu}{\sigma}} f_Z(z)dz = P\left(Z \le \frac{a-\mu}{\sigma}\right) \\
&= P(\sigma z + \mu \le a)
\end{aligned}
$$

Example:

Suppose you have $X \sim \textit{cauchy}$ (which has no defined expected value and variance):

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \qquad -\infty < x < \infty$$

Then the function:

$$\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\pi\sigma} \cdot \frac{1}{1+\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty$$

is a density despite not having a true μ or σ

2.2 Inequalities and Identities (Section 3.6)

2.2.1 Markov Inequality

Theorem

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

$$P(g(x) > r) \le \frac{E[g(x)]}{r}$$

Proof:

$$
\begin{aligned}
E[g(x)] &= \int_{-\infty}^{\infty} g(x) f_X(x)dx \\
&\ge \int g(x) f_X(x)dx \quad \{x : g(x) > r\} \\
&\ge \int r f_X(x)dx \quad \{x : g(x) > r\} \\
&= r \int f_X(x)dx \quad \{x : g(x) > r\} = r \cdot P[g(x) > r]
\end{aligned}
$$

2.2.2 Chebychev Inequality

Theorem:

Let μ and σ be *mean* and *standard deviation* for a random variable X. Consider the function $g(x) = \frac{(x-\mu)^2}{\sigma^2} > 0$.
Side Note:

$$E[g(x)] = E\left[\frac{(x-\mu)^2}{\sigma^2}\right] = \frac{1}{\sigma^2}E[(x-\mu)^2] = 1$$

Then, we have:

$$P\left(\frac{(x-\mu)^2}{\sigma^2} > r\right) \leq \frac{1}{r}$$

Let $t^2 = r$. Then, we have:

$$P\left(\frac{(x-\mu)^2}{\sigma^2} > t^2\right) \leq \frac{1}{t^2}$$

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

What the *Chebyshev Inequality* tells us is that x ($X \sim$ some distribution with mean and finite variance) can't get too far away from μ with high probability. In other words, consider the case where $t = 2$. Then we have:

$$P(|x-\mu| > 2\sigma) \leq \frac{1}{4}$$

In other words, the probability that x is more than 2 standard deviations away from μ is bounded by probability $= 0.25$. Often the probability estimates by *Chebyshev* are very conservative.

Example:

Consider $X \sim exp(1)$ with $\mu = 1$, $\sigma = 1$. By *Chebyshev inequality*:

$$\begin{aligned} P(|X-1| > 2) &= 1 - P(|X-1| \leq 2) \\ &= 1 - P(0 < X < 3) \leq \frac{1}{4} \end{aligned}$$

Exact value:

$$P(0 < X < 3) = \int_0^3 e^{-x}dx = e^{-3} \approx 0.05$$

It is **not possible** to come up with a general inequality with *tighter bounds*. The Chebychev bounds can not be improved upon without changing assumptions on X. To prove this, it is sufficient to come up with an example which has the exact bounds.

Example:

X	-1	0	1
$f(x)$	$\frac{1}{8}$	$\frac{6}{8}$	$\frac{1}{8}$

which has $E[X] = 0$ and $var(X) = \frac{1}{4}$

Taking $t = 2$ we have:

$$P\left(|X - 0| \geq 2\left(\frac{1}{2}\right)\right) = P(|X| \geq 1) = \frac{1}{4}$$

Example:

Suppose $Z \sim N(0, 1)$. Then:

$$P(|Z| > t) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t}$$

Side Note: By *Chebyshev*, we have $P(|Z| > 2) \leq \frac{1}{4}$

$$P(|Z| > 2) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-2}}{2} = 0.054$$

Proof:

$$\begin{aligned} P(|Z| > t) &= 2P(Z > t) \\ &= 2\int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\ &\leq \sqrt{\frac{2}{\pi}} \int_t^\infty \frac{z}{t} e^{-\frac{1}{2}z^2} dz \text{ , since } z > t \text{ we are multiplying by something} > 1 \\ &= \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t} \end{aligned}$$