3.3 Exponential families

Def. A family of pdf or pmf is called an **exponential family** if it has the form

$$\boxed{f(x|\theta) = c(\theta)\, h(x) \exp\left\{ \sum_{i=1}^{k} w_i(\theta) t_i(x) \right\}}, \quad x \in \mathcal{X}$$

$$= c(\theta)\, h(x) \exp \langle w(\theta), t(x) \rangle \qquad (*)$$

$w(\theta) = (w_1(\theta), \dots, w_k(\theta))$

$t(x) = (t_1(x), \dots, t_k(x))$

vector of sufficient statistics

Here

$h(x) \geq 0$

$t(x) = (t_1(x), \dots, t_k(x))$ depends on x only (data)

$w(\theta) = (w_1(\theta), \dots, w_k(\theta))$ depends on θ only (parameter)

Clearly, $c(\theta)$ is the reciprocal of the normalizing const

$$1/c(\theta) = \int_{\mathcal{X}} h(x)\, e^{\langle w(\theta), t(x) \rangle}\, dx$$

or sum, if $\mathcal{X}$ is discrete

$c(\theta)$ depends only on θ (not x).

Example 1 (Binomial) Let $X \sim \text{Binomial}(n,p)$

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0,1,\dots,n$$

$$\theta = p$$

$$= (1-p)^n \binom{n}{x} \left(p/1-p\right)^x$$

$$= \underbrace{(1-p)^n}_{c(\theta)} \underbrace{\binom{n}{x}}_{h(x)} \exp\left\{ \underbrace{x}_{t_1(x)} \underbrace{\log \frac{p}{1-p}}_{w_1(\theta)} \right\}$$

$k=1$

Example 2 (Normal) Let $X \sim N(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2\sigma^2}(x-\mu)^2, \quad x \in \mathbb{R}$$

$$\theta = (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2\sigma^2}\left(x^2 - 2x\mu + \mu^2\right)$$

$$= \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\mu^2}{2\sigma^2}}}_{c(\theta)} \exp\left\{ \underbrace{-\frac{1}{2\sigma^2}}_{w_1(\theta)} \underbrace{x^2}_{t_1(x)} + \underbrace{\frac{\mu}{\sigma^2}}_{w_2(\theta)} \underbrace{x}_{t_2(x)} \right\}$$

Other examples : Poisson, Geometric, Negative binomial, ...
Gamma, Exponential, Log-normal, ...

Not in the exponential families :
Cauchy
Mixtures

Theorem if X is a Random variable with pdf/pmf in the exponential family of the form $(*)$

then

(1) $$\mathbb{E} \sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(X) = \frac{\partial}{\partial \theta_j} \log(1/c(\theta)), \text{ for } j=1,\dots,k$$

(2) $$\mathrm{var}\left(\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(X)\right) = \frac{\partial^2}{\partial \theta_j^2} \log(1/c(\theta)) - \mathbb{E}\left(\sum_{j=1}^{k} \frac{\partial^2 w_i(\theta)}{\partial \theta_j^2} t_i(X)\right).$$

High-level statement : differentiating the logarithm of the normalizing constant with respect to a parameter results in suitable expectations of suff. statistics

Let's look at a simplified statement

$$w(\theta) = \theta = (\theta_1, \theta_2, \dots, \theta_k) \in \mathbb{R}^k$$
$$t(x) = (t_1(x), t_2(x), \dots, t_k(x)) \in \mathbb{R}^k$$

Then $f(x|\theta) = c(\theta)\, h(x) \exp\langle \theta, t(x)\rangle$

(1) gives: $\boxed{\mathbb{E}\, t_j(X) = \frac{\partial}{\partial \theta_j} \log(1/c(\theta))}$

(2) gives: $\boxed{\mathrm{var}\, t_j(X) = \frac{\partial^2}{\partial \theta_j^2} \log(1/c(\theta))}$

Proof of theorem is a simple exercise of calculus.

Example (Binomial)

$$f_X(x|p) = \underbrace{(1-p)^n}_{c(\theta)} \underbrace{\binom{n}{x}}_{h(x)} \exp\Big\{ \underbrace{x}_{t_1(x)} \underbrace{\log\frac{p}{1-p}}_{w_1(\theta)} \Big\}$$

$k=1$

$c(\theta) = (1-p)^n$ $\qquad \theta \equiv p$

$$\frac{\partial}{\partial \theta} \log\left(1/c(\theta)\right) = -\frac{\partial}{\partial p}\, n \log(1-p)$$

$$= \frac{n}{1-p}.$$

$$\frac{\partial}{\partial \theta} w_1(\theta) = \frac{\partial}{\partial p}\left(\log p - \log(1-p)\right)$$

$$= \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}$$

Applying the theorem

$$\frac{\partial}{\partial \theta} \log\left(1/c(\theta)\right) = \mathbb{E}\, \underbrace{\frac{\partial}{\partial \theta} w_1(\theta)}\ \underbrace{t_1(X)}$$

$$\frac{n}{1-p} = \mathbb{E}\, \frac{1}{p(1-p)}\, X$$

$$n = \mathbb{E}\, \frac{X}{p}$$

$$\Rightarrow \quad \mathbb{E}X = np.$$

Example (Normal)

$$f_X(x) = \frac{1}{\ } e^{-\frac{\mu^2}{2\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} x^2 + \frac{\mu}{\ } x \right\}$$

$$f_X(x) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}}_{c(\theta)} e^{-\frac{\mu^2}{2\sigma^2}} \exp\left\{ \underbrace{-\frac{1}{2\sigma^2}}_{w_1(\theta)} \underbrace{x^2}_{t_1(x)} + \underbrace{\frac{\mu}{\sigma^2}}_{w_2(\theta)} \underbrace{x}_{t_2(x)} \right\}$$

$$w(\theta) = \left(-\frac{1}{2\sigma^2}, \frac{\mu}{\sigma^2} \right)$$

$$\frac{\partial}{\partial\mu} \log(1/c(\theta)) = \frac{\partial}{\partial\mu}\left(\log\left(\sqrt{2\pi\sigma^2}\right) + \frac{\mu^2}{2\sigma^2} \right)$$

$$= \frac{\mu}{\sigma^2}$$

$$\frac{\partial}{\partial\mu} w_2(\theta) = \frac{\partial}{\partial\mu}\left(\frac{\mu}{\sigma^2}\right) = \frac{1}{\sigma^2}$$

Applying the theorem

$$\frac{\partial}{\partial\mu} \log\left(\frac{1}{c(\theta)}\right) = \mathbb{E}\left(\frac{\partial}{\partial\mu} w_2(\theta) \, t_2(X) + \underbrace{\frac{\partial}{\partial\mu} w_1(\theta)}_{0} \, t_1(X) \right)$$

$$\frac{\mu}{\sigma^2} = \frac{1}{\sigma^2} \mathbb{E}X$$

So $\mathbb{E}X = \mu$

To obtain $\mathbb{E}\,t_2(X) = \mathbb{E}X^2$:

To obtain $\mathbb{E}\, t_2(X) = \mathbb{E} X^2$:

$$\frac{\partial}{\partial\sigma} \log\left(\frac{1}{c(\theta)}\right) = \frac{\partial}{\partial\sigma}\left(\log\sqrt{2\pi\sigma^2} + \frac{\mu^2}{2\sigma^2}\right)$$

$$= \frac{1}{\sigma} - \mu^2 \frac{1}{\sigma^3}$$

$$\frac{\partial}{\partial\sigma} w_1(\theta) = \frac{\partial}{\partial\sigma}\left(-\frac{1}{2\sigma^2}\right) = \frac{1}{\sigma^3}$$

$$\frac{\partial}{\partial\sigma} w_2(\theta) = \frac{\partial}{\partial\sigma}\left(\frac{\mu}{\sigma^2}\right) = -2\mu\frac{1}{\sigma^3}$$

Applying the theorem

$$\frac{\partial}{\partial\sigma} \log\left(\frac{1}{c(\theta)}\right) = \mathbb{E}\left(\frac{\partial}{\partial\sigma} w_1(\theta)\right) t_1(X) + \left(\frac{\partial}{\partial\sigma} w_2(\theta)\right) t_2(X)$$

$$\frac{1}{\sigma} - \mu^2\frac{1}{\sigma^3} = \frac{1}{\sigma^3}\mathbb{E}X^2 - \frac{2\mu}{\sigma^3}\mathbb{E}X$$

$$\sigma^2 - \mu^2 = \mathbb{E}X^2 - 2\mu^2$$

Hence $\mathbb{E}X^2 = \sigma^2 + \mu^2$

$$\Rightarrow \operatorname{var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

$$= \mathbb{E}X^2 - \mu^2$$

$$= \sigma^2.$$