Chapter 3

Common Families of Distributions

3.4 Exponential Families

Definition 3.4.1: (Exponential Family)

A family of pmfs or pdfs is called *exponential family* if it can be expressed as

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right) \tag{3.4.1}$$

where $h(x) \geq 0$ and $t_1(x), \ldots, t_k(x)$ are real-valued functions of the observation x (they cannot depend on $\boldsymbol{\theta}$), and $c(\boldsymbol{\theta}) \geq 0$ and $w_1(\boldsymbol{\theta}), \ldots, w_k(\boldsymbol{\theta})$ are real-valued functions of the possibly vector-valued parameter $\boldsymbol{\theta}$ (they cannot depend on x).

Note: To verify that a family of pdfs or pmfs is an exponential family,

1. Identify the functions $h(x)$, $c(\boldsymbol{\theta})$, $t_i(x)$, and $w_i(\boldsymbol{\theta})$, and check that they satisfy the conditions;
2. Show that the family of pdfs or pmfs has the form of (3.4.1).

Example 3.4.1: (Examples for Exponential Families - Binomial, Poisson, Exponential, Normal Distributions)

(1) Binomial Distribution:

$$
\begin{aligned}
f(x|p) &= \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x}(1-p)^n \left(\frac{p}{1-p}\right)^x \\
&= \binom{n}{x}(1-p)^n \exp\left(x \log\left(\frac{p}{1-p}\right)\right),
\end{aligned}
$$

then

$$h(x) = \binom{n}{x},\ c(p) = (1-p)^n,\ t(x) = x \text{ and }\ w(p) = \log\left(\frac{p}{1-p}\right).$$

Note: $0 < p < 1$, and $f(x|p)$ is different for $p = 0$, $0 < p < 1$ and $p = 1$. The above formula must match all x. Therefore, $f(x|p)$ is an exponential family only if $0 < p < 1$.

(2) Poisson Distribution:

$$f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{1}{x!} e^{-\lambda} \exp\big(x \log(\lambda)\big)$$

then

$$h(x) = \frac{1}{x!},\ c(\lambda) = e^{-\lambda},\ t(x) = x \text{ and }\ w(\lambda) = \log(\lambda).$$

(3) Exponential Distribution:

$$f(x|\beta) = \frac{1}{\beta} \exp\left(-\frac{x}{\beta}\right)$$

then

$$h(x) = 1,\ c(\beta) = \frac{1}{\beta},\ t(x) = x \text{ and }\ w(\beta) = -\frac{1}{\beta}.$$

(4) Normal Distribution:

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

then

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

$$t_1(x) = -\frac{x^2}{2},\ w_1(\mu, \sigma) = \frac{1}{\sigma^2},\ t_2(x) = x \text{ and } w_2(\mu, \sigma) = \frac{\mu}{\sigma^2}.$$

Theorem 3.4.2: If X is a random variable with pdf or pmf of the form (3.4.1), then it holds for any j,

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

Remark: The theorem can be utilized as a calculational shortcut for moments of an exponential family.

Example 3.4.3: (Binomial Mean and Variance)

For Binomial Distribution, we have

$$h(x) = \binom{n}{x},\ c(p) = (1-p)^n,\ t(x) = x \text{ and } \ w(p) = \log\left(\frac{p}{1-p}\right).$$

Then,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}p}w(p) &= \frac{\mathrm{d}}{\mathrm{d}p}\log\left(\frac{p}{1-p}\right) = \frac{1}{p(1-p)},\\
\frac{\mathrm{d}^2}{\mathrm{d}p^2}w(p) &= -\frac{1}{p^2}+\frac{1}{(1-p)^2} = \frac{2p-1}{p^2(1-p)^2},\\
\frac{\mathrm{d}}{\mathrm{d}p}\log\big(c(p)\big) &= \frac{\mathrm{d}}{\mathrm{d}p}n\log(1-p) = -\frac{n}{1-p},\\
\frac{\mathrm{d}^2}{\mathrm{d}p^2}\log\big(c(p)\big) &= -\frac{n}{(1-p)^2}.
\end{aligned}$$

Therefore, from Theorem 3.4.2, we have

$$\begin{aligned}
&\mathrm{E}\left(\frac{1}{p(1-p)}X\right) = \frac{n}{1-p} \ \Rightarrow\ \mathrm{E}(X) = np,\\
&\mathrm{Var}\left(\frac{1}{p(1-p)}X\right) = \frac{n}{(1-p)^2} - \mathrm{E}\left(\frac{2p-1}{p^2(1-p)^2}X\right) \ \Rightarrow\ \mathrm{Var}(X) = np(1-p).
\end{aligned}$$

<u>Example</u>: (Normal Mean and Variance)

For Normal Distribution, we have

$$h(x) = 1,\ c(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\mu^2}{2\sigma^2}\right),$$
$$t_1(x) = -\frac{x^2}{2},\ w_1(\mu, \sigma) = \frac{1}{\sigma^2},\ t_2(x) = x \text{ and } w_2(\mu, \sigma) = \frac{\mu}{\sigma^2}.$$

Then,

$$\frac{\partial w_1(\mu, \sigma)}{\partial \mu} = \frac{\partial(1/\sigma^2)}{\partial \mu} = 0,$$
$$\frac{\partial w_2(\mu, \sigma)}{\partial \mu} = \frac{\partial(\mu/\sigma^2)}{\partial \mu} = \frac{1}{\sigma^2},$$
$$\frac{\partial w_1(\mu, \sigma)}{\partial \sigma} = \frac{\partial(1/\sigma^2)}{\partial \sigma} = -\frac{2}{\sigma^3},$$
$$\frac{\partial w_2(\mu, \sigma)}{\partial \sigma} = \frac{\partial(\mu/\sigma^2)}{\partial \sigma} = -\frac{2\mu}{\sigma^3},$$
$$\frac{\partial}{\partial \mu} \log\big(c(\mu, \sigma)\big) = \frac{\partial}{\partial \mu}\left(-\frac{\log(2\pi)}{2} - \log(\sigma) - \frac{\mu^2}{2\sigma^2}\right) = -\frac{\mu}{\sigma^2},$$
$$\frac{\partial}{\partial \sigma} \log\big(c(\mu, \sigma)\big) = \frac{\partial}{\partial \sigma}\left(-\frac{\log(2\pi)}{2} - \log(\sigma) - \frac{\mu^2}{2\sigma^2}\right) = -\frac{1}{\sigma} + \frac{\mu^2}{\sigma^3}.$$

Therefore, from Theorem 3.4.2, we have

$$\mathrm{E}\left(\frac{1}{\sigma^2}X\right) = \frac{\mu}{\sigma^2} \text{ and } \mathrm{E}\left(-\frac{2}{\sigma^3}\left(-\frac{X^2}{2}\right) - \frac{2\mu}{\sigma^3}X\right) = \frac{1}{\sigma} - \frac{\mu^2}{\sigma^3},$$

which implies

$$\mathrm{E}(X) = \mu,\ \ \mathrm{E}(X^2) = \mu^2 + \sigma^2 \ \text{ and } \ \mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}X)^2 = \sigma^2.$$

<u>Definition 3.4.5</u>: The *indicator function* of a set A, often denoted by $I_A(X)$, is the function

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

Alternatively, we can use $I(x \in A)$.

Example: Let X have a pdf given by

$$f(x|\theta) = \frac{1}{\theta} \exp\left(1 - \frac{x}{\theta}\right), \quad \text{for } \theta < x < \infty \text{ and } \theta > 0.$$

Show that this is **NOT** an exponential family. The pdf above can be written using an indicator function:

$$f(x|\theta) = \frac{1}{\theta} \exp\left(1 - \frac{x}{\theta}\right) I_{[\theta,\infty)}(x).$$

Example 3.4.1: (Exponential Families Using Indicator Functions)

(1) Binomial Distribution:

$$\begin{aligned} f(x|p) &= I_{\{0,1,\ldots,n\}}(x)\binom{n}{x}p^x(1-p)^{n-x} = I_{\{0,1,\ldots,n\}}(x)\binom{n}{x}(1-p)^n\left(\frac{p}{1-p}\right)^x \\ &= I_{\{0,1,\ldots,n\}}(x)\binom{n}{x}(1-p)^n \exp\left(x\log\left(\frac{p}{1-p}\right)\right), \end{aligned}$$

then

$$h(x) = I_{\{0,1,\ldots,n\}}(x)\binom{n}{x},\ c(p) = (1-p)^n,\ t(x) = x \text{ and } \ w(p) = \log\left(\frac{p}{1-p}\right).$$

Note: $0 < p < 1$, and $f(x|p)$ is different for $p = 0$, $0 < p < 1$ and $p = 1$. The above formula must match all x. Therefore, $f(x|p)$ is an exponential family only if $0 < p < 1$.

(2) Poisson Distribution:

$$f(x|\lambda) = I_{\{0,1,\ldots\}}(x)\frac{\lambda^x e^{-\lambda}}{x!} = I_{\{0,1,\ldots\}}(x)\frac{1}{x!}e^{-\lambda}\exp\left(x\log(\lambda)\right)$$

then

$$h(x) = I_{\{0,1,\ldots\}}(x)\frac{1}{x!},\ c(\lambda) = e^{-\lambda},\ t(x) = x \text{ and } \ w(\lambda) = \log(\lambda).$$

(3) Exponential Distribution:

$$f(x|\beta) = I_{[0,\infty)}(x)\frac{1}{\beta}\exp\left(-\frac{x}{\beta}\right)$$

then

$$h(x) = I_{[0,\infty)}(x),\ c(\beta) = \frac{1}{\beta},\ t(x) = x \text{ and } \ w(\beta) = -\frac{1}{\beta}.$$

(4) Normal Distribution:

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

then

$$h(x) = 1,\ c(\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\mu^2}{2\sigma^2}\right),$$
$$t_1(x) = -\frac{x^2}{2},\ w_1(\mu,\sigma) = \frac{1}{\sigma^2},\ t_2(x) = x \text{ and } w_2(\mu,\sigma) = \frac{\mu}{\sigma^2}.$$

Definition: (Reparameterization of Exponential Families)

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

where $\boldsymbol{\eta} = (\eta_1,\dots,\eta_k)$, $\eta_i = w_i(\boldsymbol{\theta})$ are *natural parameters*, $h(x)$ and $t_i(x)$ are the same as in the original parameterization, and

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

to ensure that the pdf integrates to 1. The set

$$\mathcal{H} = \left\{\boldsymbol{\eta} = (\eta_1,\dots,\eta_k) : \int_{-\infty}^{\infty} h(x)\exp\left(\sum_{i=1}^{k}\eta_i t_i(x)\right)dx < \infty\right\}$$

is called the *natural parameter space* for the family. (The integral is replaced by a sum over the values of x for which $h(x) > 0$ if X is discrete.) Since the original $f(x|\boldsymbol{\theta})$ in (3.4.1) is a pdf or pmf, it must hold that

$$\left\{\boldsymbol{\eta} = \big(w_1(\boldsymbol{\theta}),\dots,w_k(\boldsymbol{\theta})\big) : \boldsymbol{\theta}\in\Theta\right\} \subset \mathcal{H}.$$

Example 3.4.6: (Reparameterization of Normal Distribution)

For Normal distribution, we have

$h(x) = 1,\ c(\mu, \sigma^2) = \dfrac{1}{\sqrt{2\pi}\sigma} \exp\left(-\dfrac{\mu^2}{2\sigma^2}\right),$

$t_1(x) = -\dfrac{x^2}{2},\ w_1(\mu, \sigma^2) = \dfrac{1}{\sigma^2},\ t_2(x) = x$ and $w_2(\mu, \sigma^2) = \dfrac{\mu}{\sigma^2}.$

Let $\eta_1 = w_1(\mu, \sigma^2) = 1/\sigma^2$ and $\eta_2 = w_2(\mu, \sigma^2) = \mu/\sigma^2$. The Normal distribution can be reparameterized as:

$$f(x|\eta_1, \eta_2) = \frac{\sqrt{\eta_1}}{\sqrt{2\pi}} \exp\left(-\frac{\eta_2^2}{2\eta_1}\right) \exp\left(-\frac{\eta_1}{2}x^2 + \eta_2 x\right),$$

where $\eta_1 = 1/\sigma^2$ and $\eta_2 = \mu/\sigma^2$. The natural parameter space is that $\eta_1 > 0$ and $-\infty < \eta_2 < \infty$.

Definition 3.4.7: A *curved exponential family* is a family of densities of the form (3.4.1) for which the dimension of the vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$ is equal to $d < k$. If $d = k$, the family is a *full exponential family.*

Example 3.4.8: Normal distribution with mean μ and variance $\sigma^2 = \mu^2$.

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

Let $\eta_1 = 1/\mu^2$ and $\eta_2 = 1/\mu$. The Normal distribution $\mathrm{n}(\mu, \mu^2)$ can reparameterized as:

$$f(x|\eta_1, \eta_2) = \frac{\sqrt{\eta_1}}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\right) \exp\left(-\frac{\eta_1}{2}x^2 + \eta_2 x\right).$$

Since $d = 1$ and $k = 2$, it is a curved exponential family.

Example 3.4.9: (Normal Approximation)

$X_1, \ldots, X_n$ are sampled from a Poisson(λ) population, then the distribution of $\bar{X} = \sum_{i=1}^{n} X_n$ is approximately (according to the Central Limit Theorem)

$$\bar{X} \sim \mathrm{n}(\lambda, \lambda/n),$$

which is a curved exponential family.

$X_1, \ldots, X_n$ are iid Bernoulli(p), then the distribution of $\bar{X}$ is approximately

$$\bar{X} \sim \mathrm{n}\big(p, p(1-p)/n\big),$$

which is also a curved exponential family.

Remark:

1. Theorem 3.4.2 also applied to curved exponential families.
2. Exponential families have nice properties that are very useful in statistical inference.

3.5 Location and Scale Families

Three types of families of interest:

1. Location Families
2. Scale Families
3. Location-Scale Families

Note:

1. Each of these families is constructed from a single pdf (or pmf) known as the standard pdf (pmf) for the family;
2. All other pdfs (or pmfs) in the family are obtained by transforming the standard pdf (or pmf) in a prescribed way.

Theorem 3.5.1: Let $f(x)$ be any pdf and let μ and $\sigma > 0$ be any given constants. Then the function

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

is a valid pdf.

Proof.

$$g(x|\mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) \geq 0$$

$$\int_{-\infty}^{\infty} g(x|\mu, \sigma)dx = \int_{-\infty}^{\infty} \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) dx \overset{\left(y=\frac{x-\mu}{\sigma}\right)}{=\!=\!=} \int_{-\infty}^{\infty} f(y)dy = 1.$$

■

Definition 3.5.2: Let $f(x)$ be any pdf. Then the family of pdfs $f(x-\mu)$, indexed by the parameter μ $(-\infty < \mu < \infty)$, is called the *location family* with standard pdf $f(x)$ and μ is called the location parameter for the family.

Remark:

1. The effect of location parameters shifts the density to the left or right but the shape remains unchanged.
2. If Z has a pdf $f(z)$, then $X = Z + \mu$ has density $f(x-\mu)$.

Example 3.5.3: (Exponential Location Family)

Let

$$f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

To form a location family, we replace x with $x - \mu$ to obtain

$$f(x|\mu) = \begin{cases} e^{-(x-\mu)} & x-\mu \geq 0 \\ 0 & x-\mu < 0 \end{cases} = \begin{cases} e^{-(x-\mu)} & x \geq \mu \\ 0 & x < \mu \end{cases}.$$

If we use the indicator function to express this, we have

$$f(x|\mu) = e^{-(x-\mu)} I_{[0,\infty)}(x-\mu) = e^{-(x-\mu)} I_{[\mu,\infty)}(x).$$

Definition 3.5.4: Let $f(x)$ be any pdf. Then for any $\sigma > 0$, the family of pdfs $\frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$, indexed by the parameter σ, is called the *scale family* with standard pdf $f(x)$ and σ is called the scale parameter of the family.

Remark: The effect of scale parameter σ is either to stretch or to contract the graph $f(x)$ maintaining the same basic shape of the graph.

Example: (Normal Distribution)

$$f(x|\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right), \quad -\infty < x < \infty, \sigma > 0,$$

where σ is the scale parameter of the scale family with standard pdf below

$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad -\infty < x < \infty.$$

Definition 3.5.5: Let $f(x)$ be any pdf. Then for any μ $(-\infty < \mu < \infty)$, and any $\sigma > 0$, the family of pdfs $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, indexed by the parameter (μ, σ), is called the *location-scale family* with standard pdf $f(x)$; μ is called the location parameter and σ is called the scale parameter.

Example: (Normal and Double Exponential Distributions)

$$f(x|\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0.$$

$$f(x|\sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|x-\mu|}{\sigma}\right), \quad -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0.$$

Theorem 3.5.6: Let $f(\cdot)$ be any pdf. Let μ be any real number, and let σ be any positive real number. Then X is a random variable with pdf $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$ if and only if there exists a random variable Z with pdf $f(z)$ and $X = \sigma Z + \mu$.

Proof. To prove the "if" part, define $g(z) = \sigma z + \mu$. Then $X = g(Z)$, g is a monotone function,

$$g^{-1}(x) = \frac{x-\mu}{\sigma} \quad \text{and} \quad \left|\frac{\mathrm{d}}{\mathrm{d}x} g^{-1}(x)\right| = \frac{1}{\sigma}.$$

Thus by Theorem 2.1.5, the pdf of X is

$$f_X(x) = f_Z\big(g^{-1}(x)\big) \left|\frac{\mathrm{d}}{\mathrm{d}x} g^{-1}(x)\right| = \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right).$$

It is similar to prove the "only if" part: define $g(x) = (x-\mu)/\sigma$ and let $Z = g(X)$. ■

Theorem 3.5.7: Let Z be a random variable with pdf $f(z)$. Suppose $\mathrm{E}Z$ and $\mathrm{Var}Z$ exist. If X is a random variable with pdf $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, then

$$\mathrm{E}X = \sigma \mathrm{E}Z + \mu \quad \text{and} \quad \mathrm{Var}X = \sigma^2 \mathrm{Var}Z.$$

Proof. Based on Theorem 3.5.6, we have $X = \sigma Z + \mu$. ■