Exponential Families

- Turns out all the classical (univariate) distributions that we have covered can be written as special cases of a more general construction.

$\downarrow$

$\boxed{\text{Exponential Family}}$

Example:

$\rightarrow$ Poisson pdf

$$f_\mu(x) = \frac{e^{-\mu}\mu^x}{x!}$$

Consider another Poisson pdf, with fixed parameter μ_0

$$f_{\mu_0}(x) = \frac{e^{-\mu_0}\mu_0^x}{x!}$$

If we divide them, we get

$$\frac{f_\mu(x)}{f_{\mu_0}(x)} = e^{-(\mu-\mu_0)}\left(\frac{\mu}{\mu_0}\right)^x$$

$$\rightarrow f_\mu(x) = e^{-(\mu-\mu_0)}\left(\frac{\mu}{\mu_0}\right)^x \cdot f_{\mu_0}(x)$$

$\rightarrow$ we're writing Poi(μ) pdf in terms of the Poi(μ_0) pdf. we're "tilting" $f_{\mu_0}(x)$ by

We can rewrite this as

$$f_\mu(x) = e^{\alpha x - \psi(\alpha)} \cdot f_{\mu_0}(x)$$

where $\alpha = \log\left(\frac{\mu}{\mu_0}\right)$ and $\psi(\alpha) = \mu_0(e^\alpha - 1)$

$$\exp\left\{\log\left(\frac{\mu}{\mu_0}\right)\cdot x - \mu_0\left(e^{\log(\mu/\mu_0)} - 1\right)\right\}$$

$$\exp\left\{\log\left[\left(\frac{\mu}{\mu_0}\right)^x\right] - \mu_0\left(\frac{\mu}{\mu_0} - 1\right)\right\}$$

$$\exp\left\{\log\left[\left(\frac{\mu}{\mu_0}\right)^x\right] - (\mu - \mu_0)\right\} \checkmark$$

Then, we can take any one Poisson Distribution $f_{\mu_0}(x)$ and for any value $\mu > 0$, let $\alpha = \log\left(\frac{\mu}{\mu_0}\right)$ and calculate

Exponential tilt

$$\tilde{f}_\mu(x) = e^{\alpha x} f_{\mu_0}(x)$$

and divide by $e^{\psi(\alpha)}$ to get Poisson density $f_\mu(x)$.

$\psi(\cdot)$ not a function of x.

Then $e^{\psi(\alpha)}$ is your renormalization constant.

In the same way that we write the Poisson pmf in exponential family form, we can write other distributions with more than 1 parameter as exponential families.

$$f_\alpha(x) = e^{\alpha^T y - \psi(\alpha)} \cdot f_0(x) \quad \text{for } \boxed{\begin{array}{l}\alpha \in A \\ A \subseteq \mathbb{R}^p\end{array}}$$

"α is a p-dimensional vector of parameters".

$y = t(x)$: sufficient statistic vector.

The normalizing function $e^{\psi(\alpha)}$

$$e^{\psi(\alpha)} = \int_{\mathcal{X}} e^{\alpha^T y} f_0(x)\, dx$$

"$e^{\psi(\alpha)}$ is the value of this integral, and we must divide over it for the pdf to integrate to 1 (or the pmf to sum to 1)".

For instance in the previous Poisson example, we got

$$e^{\psi(\alpha)} = \sum_{0}^{\infty} e^{\alpha x} f_{\mu_0}(x).$$

because x is discrete.

A 2-parameter example would be the Gamma distribution

$$f(x) = \frac{x^{\nu-1} e^{-x/\sigma}}{\sigma^{\nu} \Gamma(\nu)}$$

$\sigma > 0$, $\nu > 0$

$\alpha = (\alpha_1, \alpha_2) = \left(-\frac{1}{\sigma}, \nu\right)$

$(y_1, y_2) = (x, \log x)$

and $\psi(\alpha) = \nu \log \sigma + \log \Gamma(\nu)$

$= -\alpha_2 \log\{-\alpha_1\} + \log\{\Gamma(\alpha_2)\}$

and do the algebra yourself...

→ Just like we can write 1 or 2-dim classical families in exponential family form, we can extend the dimensionality of $\underline{\alpha}$ to get more general/flexible parametric distributions

Having said this, under repeated sampling, we can take the exponential family pdf

$$f_\alpha(x) = e^{\alpha^T y - \psi(\alpha)} \cdot f_0(x_i) \quad \text{(density)}$$

$$\downarrow$$

$$f_\alpha(\underline{X}) = \prod_{i=1}^{n} e^{\alpha^T y_i - \psi(\alpha)} f_0(x_i) \quad \text{(likelihood)}$$

$$= e^{\alpha^T \sum_{i=1}^{n} y_i - n\psi(\alpha)} f_0(x_i)$$

recall $\underline{y} = t(\underline{x})$

$$\boxed{= e^{n(\alpha^T \bar{y} - \psi(\alpha))} f_0(x_i)}$$

→ So then we have that once we have defined $\underline{y} = t(\underline{x})$, we only need its mean $\bar{y}$ to obtain its likelihood.

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Argument 1	Argument 2
You can write many of the classical pdfs/pmfs as exponential families. → need clever substitutions, transformations, etc.	→ you can come up with almost any $t(\underline{x})$ you would like and use the exp. family likelihood to fit your data.

Example of how to use the exponential family construction to fit data.

- Use gfr data.

gfr

x_1

x_2

x_3

$\vdots$

x_n

① Define our sufficient statistic vector

$$y(x) = (x, x^2, x^3, \ldots, x^7)$$

(Apply to each observation).

② This will allow us to fit a 7-parameter exponential family!