

Exponential families of distributions

X r.v. has exp. family distribution if it has a d.f. F lies in the set

$$\{F_{\theta} : \theta \in \Omega\}$$

with densities

$$(1) \quad p_{\theta}(x) = \exp\left(\sum_{i=1}^J \eta_i(\theta) T_i(x) - B(\theta)\right) h(x)$$

w.r.t. some measure μ .

For us this means one of two things

(i) X discrete r.v., $p_{\theta}(x)$ is a probability mass function, μ counting measure

$$\int_A f(x) d\mu(x) = \sum_{i \in A} f(i)$$

(ii) X continuous r.v., $p_{\theta}(x)$ is a probability density function, μ Lebesgue measure and

$$\int_A f(x) d\mu(x) = \int_A f(x) dx$$

(1)

θ parameter, $\eta_i(\theta)$ functions of θ ,

$T_i(x)$ a statistic

$B(\theta)$ "normalising function"

$h(x)$ " " " " "

A more convenient form

$$(2) \quad p_{\eta}(x) = \exp\left(\sum_{i=1}^J \eta_i T_i(x) - A(\eta)\right) h(x)$$

canonical form of exp. fam. distribution.

why is this interesting

a) a generalisation of many distributions we are familiar with.

b) ~~the~~ the form (2) is mathematically nice, it gives us a lot of info

(i) $A(\eta)$ can be used to derive all moments of \bar{X} , ($E(\bar{X}), E(\bar{X}^2), \dots$)

(ii) any estimator (reasonable) can be based on the statistics $T_1(x), \dots, T_g(x)$. (2)

We have that (2) is ~~assumed~~ supposed to define a density

$$\begin{aligned} 1 &= \int p_{\eta}(x) d\mu(x) \\ &= \int \exp\left(\sum_{i=1}^J \eta_i T_i(x)\right) h(x) d\mu(x) \\ &\quad \cdot e^{-A(\eta)} \end{aligned}$$

The set of points (η_1, \dots, η_J) for which

$$\int \exp\left(\sum_{i=1}^J \eta_i T_i(x)\right) h(x) d\mu(x) < \infty$$

is called the natural parameter space,

(H).

EX:

$\bar{X} \in N(\bar{\mu}, \sigma^2)$. Then the p.d.f. is

$$p_{\theta}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\bar{\mu})^2/2\sigma^2}$$

with $\theta = (\bar{\mu}, \sigma^2)$.

(3)

We see that, expand the square,

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\underbrace{\frac{\xi}{\sigma^2} x}_{\eta_1} - \underbrace{\frac{1}{2\sigma^2} x^2}_{T_1(x)} - \underbrace{\frac{\xi^2}{2\sigma^2}}_{T_2(x)}\right)$$

η_2

The natural parameters are

$$(\eta_1, \eta_2) = \left(\frac{\xi}{\sigma^2}, -\frac{1}{2\sigma^2}\right)$$

2-parameter exp. family

Natural parameter space is $\mathbb{R} \times \mathbb{R}_-$.

#

Sometimes the parameters in (2) or (1) can be seen to be linearly dependent. Then we can reduce the dimension of the parameter space.

If we do not do that, we have an identifiability problem.

(9)

Def. If $\theta_1 \neq \theta_2$ & imply
 $P_{\theta_1} = P_{\theta_2}$

then the parameter is said to be non-identifiable.

EX (multinomial distribution)

Do n independent experiments, $s+1$ different possible outcomes, each with probability

$$p_i = P(\text{outcome of type "i"}) \\ i = 0, \dots, s.$$

$\mathbf{X} =$

(X_0, X_1, \dots, X_s) data observation

$X_i =$ number of ~~out~~ outcomes
of type i , $i = 0, \dots, s$.

The random vector \mathbf{X} has density w.r.t. counting measure, or probability density function

$$x = (x_0, x_1, \dots, x_s) \in \mathbb{R}^{s+1}$$

$$p(x) = p(x_0, x_1, \dots, x_s)$$

$$= \frac{n!}{x_0! \dots x_s!} p_0^{x_0} p_1^{x_1} \dots p_s^{x_s}$$

But $x_0 + \dots + x_s = n$, so this can be rewritten

$$\exp\left(\underbrace{-A(\eta)}_{n \log p_0} + x_1 \log \frac{p_1}{p_0} + \dots + x_s \log \frac{p_s}{p_0}\right) \cdot h(x)$$

This is an s -dim exponential family, with $\eta_1 = \log \frac{p_1}{p_0}, \dots, \eta_s = \log \frac{p_s}{p_0}$ ~~natural~~ canonical parameters, the natural parameter space is \mathbb{R}^s .

#

(6)

If in the representation (2) neither T_i 's nor η_i 's ~~are~~ are linearly dependent and the natural parameter space contains an r -dim rectangle, then the model is called full rank exponential fam. distr. model.

Ex. curved exponential family

$X \in N(\xi, \sigma^2)$ and $\xi = \sigma$, then

$$p_{\theta}(x) = \frac{1}{\sqrt{2\pi\xi^2}} \exp\left(\frac{1}{\xi}x - \frac{1}{2\xi^2}x^2 - \frac{1}{2}\right)$$

Formerly 2-dim exp family $T_1(x), T_2(x) = x^2,$
 $= x$

by but

$$(\eta_1, \eta_2) = \left(\frac{1}{\xi}, -\frac{1}{2\xi^2}\right)$$

2-dim parameter lies on a curve in \mathbb{R}^2 .
 \neq

If the natural parameters depend on each other in a non-linear way we have a curved exponential family.

If we have two indep. r.v. X, Y
with k -dim exponential family distributions

$$\exp\left(\sum \eta_i T_i(x) - A(\eta)\right) h(x) = p_{\eta}^{(x)}$$

$$\exp\left(\sum \eta_i U_i(y) - C(\eta)\right) k(y) = p_{\eta}^{(y)}$$

w.r.t. measures μ, ν , then the joint
distribution of (X, Y) is also k -dim
with

$$\exp\left(\sum_{i=1}^k \eta_i (T_i(x) + U_i(y)) - (A(\eta) + C(\eta))\right) \cdot h(x)k(y)$$

In particular if X_1, \dots, X_n are i.i.d.
with exp. family as in (2) then
they have joint density

$$\exp\left(\sum_{i=1}^k \eta_i T_i'(x) - nA(\eta)\right) h_1(x_1) \dots h_n(x_n)$$

where $T_i'(x) = \sum_{j=1}^n T_i(x_j)$.

(8)

Ex (read on your own)

X_1, \dots, X_n i.i.d. $N(\xi, \sigma^2)$. The joint density of (X_1, \dots, X_n) is

$$\exp\left(\frac{\xi}{\sigma^2} \sum x_i - \frac{1}{2\sigma^2} \sum x_i^2 - \frac{n}{2\sigma^2} \xi^2\right) \cdot \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n.$$

This is a 2-dim exp. family.

$$T_1(x) = \sum x_i, \quad T_2(x) = \sum x_i^2.$$

#

Theorem.

Suppose f integrable w.r.t. μ , and η is a point in the interior of the natural parameter space (η) . Then

$$g(\eta) = \int f(x) \exp\left(\sum_{i=1}^k \eta_i T_i(x)\right) h(x) d\mu(x)$$

is cont. and has derivatives of all order, w.r.t. μ . They can be obtained by differentiating inside the integral

#

(9)

We use this # to derive the moments of the distribution.

Look at the identity

$$1 = \int \underbrace{\exp(\sum \eta_i T_i(x) - A(\eta)) h(x)}_{p_\eta(x)} d\mu(x)$$

Differentiate w.r.t η_j

$$0 = \int \left(T_j(x) - \frac{\partial}{\partial \eta_j} A(\eta) \right) \exp(\sum \eta_i T_i(x) - A(\eta)) h(x) d\mu(x)$$

$$0 = \int T_j(x) p_\eta(x) d\mu(x) - \int \frac{\partial}{\partial \eta_j} A(\eta) p_\eta(x) d\mu(x)$$

$$0 = E(T_j(X)) - \frac{\partial}{\partial \eta_j} A(\eta)$$

\Rightarrow

$$E(T_j(X)) = \frac{\partial}{\partial \eta_j} A(\eta).$$

Continue with the argument to obtain e.g.

$$E(T_j(X) T_k(X)) = \frac{\partial^2}{\partial \eta_j \partial \eta_k} A(\eta).$$

We can define all moments

$$\alpha_{r_1, \dots, r_s} = \bar{E}(T_1^{r_1} \dots T_s^{r_s})$$

and, if it exists, the moment-generating function

$$\begin{aligned} M_T(u_1, \dots, u_s) &= \sum_{\mathbf{r}} M_{T_1, T_2, \dots, T_s}(u_1, \dots, u_s) \\ &= \bar{E}(e^{u_1 T_1 + \dots + u_s T_s}) \end{aligned} \quad (1)$$

If it exists in a neighbourhood $\|u\| \leq \delta$ around 0, then we can make a Taylor expansion of (1) around 0

$$M_T(u_1, \dots, u_s) = \sum \alpha_{r_1, \dots, r_s} \frac{u_1^{r_1} \dots u_s^{r_s}}{r_1! \dots r_s!}$$

Therefore

If X exp. family distributed ~~as~~ ω
in (2) then

$$M_T(u) = \frac{e^{A(\eta+u)}}{e^{A(\eta)}}$$

#