Example 5 *Let $U = (U_1, \ldots, U_n)$ be a random vector with a distribution function F on $\mathbb{R}^n$. Then for $a \in \mathbb{R}, b \in \mathbb{R}^+$,*

$$\begin{aligned} U + a &= (U_1 + a. \ldots, U_n + a), \\ bU &= (bU_1, \ldots, bU_n), \\ a + bU &= (a + bU_1, \ldots, a + bU_n), \end{aligned}$$

generate location, scale and location/scale families, respectively. (This follows by checking that the corresponding distributions for $U + a, bU$ and $a + bU$ satisfy the definitions for location, scale and location/scale families). □

Example 6 *Let $U = (U_1, \ldots, U_p)$ be a random vector, with U_i independent $N(0, 1)$. Let $a' = (a_1, \ldots, a_p) \in \mathbb{R}^p$ and B an invertible $p \times p$ matrix be arbitrary, and define*

$$\begin{pmatrix} X_1 \\ .. \\ X_p \end{pmatrix} = \begin{pmatrix} a_1 \\ .. \\ a_p \end{pmatrix} + B \begin{pmatrix} U_1 \\ .. \\ U_p \end{pmatrix},$$

or, using matrix formulation, as

$$X = a + BU.$$

Then X is a Gaussian vector and

$$\begin{aligned} E(X) &= a. \\ \mathrm{Cov}(X) &= E((X - a)(X - a)') \\ &= E(BUU'B') \\ &= BB' \\ &=: \Sigma \end{aligned}$$

Thus the distribution of X lies in is the family of non singular p-variate Normal distributions: If

$$f_U(u) = \frac{1}{(2\pi)^{p/2}} e^{-u'u/2},$$

then

$$f_X(x) = \frac{1}{(2\pi)^{p/2}|B|} e^{-(x-a)'\Sigma^{-1}(x-a)},$$

by the formula for variable transformation in integrands. □

Example 7 *(Linear model in Normal distribution) Assume $U = (U_1, \dots, U_n)$ is a stochastic vector and let*

$$X_i = a_i + bU_i,$$

where $b > 0$ and $a = (a_1, \dots, a_n)$ lying in an s-dimensional subspace $\Omega \subset \mathbb{R}^n$ (meaning that every vector a can be written as

$$a_i = \sum_{j=1}^{s} d_{ij}\beta_j,$$

with $(\beta_1, \dots, \beta_s) \in \mathbb{R}^s$ and $D = (d_ij)$ an $n \times s$ matrix of rank s). Then if $(U_1, \dots, U_n)$ are i.i.d. $N(0,1)$

$$f_X(x) = \frac{1}{(\sqrt{2\pi}b)^n} e^{-\sum (x_i - a_i)^2/2b^2}.$$

□

Example 8 *(Nonparametric family with support on $\mathbb{R}$) Assume $U_1, \dots, U_n$ are i.i.d. r.v.'s with distribution $F = N(0,1)$. Let*

$$\mathcal{G} = \{g : \; g : \mathbb{R} \to \mathbb{R} \text{ continuous and strictly increasing}, \lim_{u \to \pm} g(u) = \pm\infty\}.$$

Then under the binary operation $\cdot$ equal to composition,

$$g_1 \cdot g_2(x) = g_1(g_2(x))$$

G is a group (thus (i) $\cdot$ is a map $\mathcal{G} \times \mathcal{G} \to \mathcal{G}$, (ii) there is a unit e in $\mathcal{G}$ and (iii) for every g there is an inverse g^{-1} in $\mathcal{G}$ such that $e = g \cdot g^{-1}$). The group family of distributions

$$\mathcal{P} = \{F(g^{-1}) : g \in \mathcal{G}\},$$

is the the set of all continuous distributions with support on $\mathbb{R}$ (the corresponding r.v.'s are given by $X_i = g(U_i)$). □

Example 9 *(Symmetric distributions) Let $F = N(0,1)$, and define*

$$\mathcal{G} = \{g : \; g : \mathbb{R} \to \mathbb{R} \text{ continuous and strictly increasing}, \\ g(-u) = -g(u), \lim_{u \to \pm} g(u) = \pm\infty\}.$$

Then $\mathcal{G}$ is group under composition and

$$\mathcal{P} = \{F(g^{-1}) : g \in G\}$$

is the class of distributions with support on $\mathbb{R}$ that are symmetric around the origin. □

3.3 Exponential families

Let $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$ be a family of distributions. Assume that P_θ has a density $p_\theta = dP_\theta/d\mu$, with respect to the ($\sigma$-finite) measure μ, that is of the form

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

Then the set $\mathcal{P}$ is called an $s-$dimensional exponential family. Here $\{\eta_i\}, B$ are real-valued functions defined on the parameter space Ω, and $\{T_i\}$ are statistics.

An alternative form is the so called canonical form

$$p_\eta(x) \quad = \quad \exp(\sum_{i=1}^{s} \eta_i T_i(x) - A(\eta))h(x),$$

with $\eta = (\eta_1 \ldots, \eta_s)$ the so called canonical parameters.

Note that in our case μ is either Lebesgue measure, in which case p_θ (and p_η) is the density function for a continuous r.v., or μ is counting measure, in which case p_θ is the probability mass function of a discrete r.v..The function h is used to prevent the use of more elaborate measures μ than Lebesgue measure and counting measure.

We note first that since the p_η are supposed to be densities corresponding to a probability measure they should integrate to one, i.e.

$$\int p_\eta(x)\, d\mu(x) \quad = \quad \int \exp\left(\sum \eta_i T_i(x)\right) h(x)\, d\mu(x) e^{-A(\eta)} = 1,$$

which is possible if and only if

$$\int \exp\left(\sum \eta_i T_i(x)\right) h(x)\, d\mu(x) \quad < \quad \infty.$$

If this holds, the normalizing function $A(\eta)$ can be chosen as

$$A(\eta) \quad = \quad \log\left(\int \exp\left(\sum \eta_i T_i(x)\right) h(x)\, d\mu(x)\right).$$

Definition 2 *The set* $\Omega = \{\eta = (\eta_1, \ldots, \eta_s)\}$ *for which*

$$\int \exp\left(\sum \eta_i T_i(x)\right) h(x)\, d\mu(x) \quad < \quad \infty.$$

holds is called the natural parameter space. □

Example 10 *Assume* X *is distributed as* $N(\xi, \sigma^2)$*, so that* $\theta = (\xi, \sigma^2)$*. Then the density (w.r.t. Lebesgue measure on* $\mathbb{R}$*) is*

$$p_\theta(x) \quad = \quad \frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{\xi}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \frac{\xi^2}{2\sigma^2}).$$

Therefore, the natural parameters are $(\frac{\xi}{\sigma^2}, \frac{-1}{2\sigma^2})$*, and the natural parameter space is* $\mathbb{R} \times (-\infty, 0]$. □

If the statistics $T_1, \dots, T_s$ are linearly dependent, the number of terms in the expression can be reduced (until they no longer are linearly dependent). If this reduction is not performed, the parameters (or equivalently the probability measures P_θ) will not be identifiable.

Definition 3 *Let $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$ with Ω some parameter space. If there are values $\theta_1 \neq \theta_2$ such that $P_{\theta_1} = P_{\theta_2}$, the family of distributions $\mathcal{P}$ is said to be unidentifiable for θ.* □

If $\mathcal{P}$ is nonidentifiable for the parameter θ, one can not from knowledge of the particular $P_\theta \in \mathcal{P}$ that is the true (unknown) distribution draw conclusion about which $\theta \in \Omega$ that is the true (unknown) parameter. When one wants to draw inference from an observable x of $X \sim P_\theta$ one can at best say which distribution is the true distribution, and the knowledge of this then does not lead to the true parameter θ. This situation is therefore undesirable.

Example 11 *(Multinomial distribution) Assume we make n independent trials with $s+1$ different possible outcomes, $O_0, \dots, O_s$ and let*

$$p_i = P(\text{outcome of type } O_i), i = 0, 1, \dots, s.$$

Let X_i be the number of outcomes of type i

$$X_i = \sum_{i=1}^{n} 1\{\text{outcome } O_i \text{ in trial } i\},$$

for $i = 1, \dots, s$. Then, with $\theta = (p_0, \dots p_s)$ and $x = (x_0, \dots, x_s)$,

$$\begin{aligned} p_\theta(x) &= P_\theta(X_0 = x_0, \dots, X_s = x_s) \\ &= \frac{n!}{x_0! \cdot \ldots \cdot x_s!} p_0^{x_0} \cdot \ldots \cdot p_s^{x_s} \\ &= \exp(x_0 \log x_0 + \ldots + x_s \log x_s) h(x), \end{aligned}$$

with $h(x) = n!/(x_0! \cdot \ldots \cdot x_s!)$. But, since $x_0 + \ldots + x_s = n$, this can be written as

$$\begin{aligned} p_\theta(x) &= \exp\left(x_0 \log x_0 + x_1(\log x_1 - \log x_0) + x_1 \log x_0 + \right. \\ &\quad \left. \ldots + x_s(\log x_s - \log x_0) + x_s \log x_0\right) h(x) \\ &= \exp(n \log p_0 + x_1 \log(p_1/p_0) + \ldots + x_s \log(p_s/p_0)) h(x), \end{aligned}$$

This is an s-dimensional exponential family, with natural parameters $\eta_i = \log(p_i/p_0)$, $i = 1, \dots, s$, $A(\eta) = -n \log(p_0)$, and natural parameter space $\mathbb{R}^s$. □

Clearly also linear dependence between the natural parameters η gives rise to nonidentifiability problems.

Definition 4 *If neither the T_i nor the η_i are linearly dependent the canonical representation is said to be minimal.* □