k-parameter exponential family

A family of distributions, $\{P_\theta, \theta \in \Theta\}$ with density $f(\mathbf{x}|\theta)$ is called a k-parameter exponential family if there exist real-valued functions $c_1(\theta), \ldots, c_k(\theta)$ and $d(\theta)$, real-valued functions $T_1(\mathbf{x}), \ldots, T_k(\mathbf{x})$ and $S(\mathbf{x})$ on $\mathcal{R}^n$, and $A \subset \mathcal{R}^n$ such that

$$f(\mathbf{x}|\theta) = \left\{ \exp\left(\sum_{j=1}^{k} c_j(\theta) T_j(\mathbf{x}) + d(\theta) + S(\mathbf{x}) \right) \right\} I_A(\mathbf{x}).$$

By the Factorization Theorem, $(T_1(\mathbf{X}), \ldots, T_k(\mathbf{X}))$ is sufficient for θ. Note that, in a k-parameter exponential family, $(T_1, \ldots, T_k)$ is the k-dimensional sufficient statistics for θ. The parameter here is θ, and not $(c_1(\theta), \ldots, c_k(\theta))$.

Example. $X \sim N(\mu, \sigma^2)$. Then

$$
\begin{aligned}
f(x|\mu, \sigma^2) &= (2\pi)^{-1/2} \sigma^{-1} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right) I_{(-\infty, \infty)}(x) \\
&= (2\pi)^{-1/2} \sigma^{-1} \exp\left(-\frac{x^2}{2\sigma^2} + \frac{\mu}{\sigma^2} x - \frac{\mu^2}{2\sigma^2} \right) I_{(-\infty, \infty)}(x) \\
&= \exp\left(\frac{\mu}{\sigma^2} x - \frac{1}{2\sigma^2} x^2 + -\frac{\mu^2}{2\sigma^2} - \log(\sigma) - \frac{1}{2}\log(2\pi) \right) I_{(-\infty, \infty)}(x).
\end{aligned}
$$

We can take $T_1(x) = x$, $T_2(x) = x^2$, $c_1(\mu, \sigma^2) = \frac{\mu}{\sigma^2}$, $c_2(\mu, \sigma^2) = -\frac{1}{2\sigma^2}$, $d(\mu, \sigma^2) = -\log(\sigma) - \frac{\mu^2}{2\sigma^2}$, $S(x) = -\frac{1}{2}\log(2\pi)$, $A = \mathcal{R}$ to see that it is a 2-parameter exponential family. Now consider $X_1, \ldots, X_m$ i.i.d from $N(\mu, \sigma^2)$. Then $(\sum_{i=1}^{m} X_i, \sum_{i=1}^{m} X_i^2)$ is sufficient for (μ, σ^2).

Note that in a k-parameter exponential family, θ need not be k-dimensional. For example, consider $N(\theta, \theta^2)$, which is a 2-parameter exponential family, but the parameter is $\theta \in \mathcal{R}^1$.

Ancillary Statistics

There are various results in classical statistics that show a sufficient statistic contains all the information about θ in the data $\mathbf{X}$. At the other end is a statistic whose distribution does not depend on θ and so contains no information about θ. Such a statistic is called *ancillary*.

Definition. Let $\mathbf{X} \sim P_\theta$. A statistic $S(\mathbf{X})$ whose distribution does not depend on the parameter θ is called an ancillary statistic.

Alone, ancillary statistic contains no information about the parameter. However, combination of ancillaries may be informative. For example, consider

(X, Y) which is bivariate normal, with both means equal to 0, both variances equal to 1, and covariance of ρ. Then both X and Y are ancillary by themselves, but together they are informative about ρ. Ancillary statistics are easy to exhibit if $X_1, \ldots, X_n$ are i.i.d. with a location-scale family of densities.

Example. $X_1, \ldots, X_n$ are i.i.d. $N(\theta, 1)$, $-\infty < \theta < \infty$. Then $\bar{X}$ is minimal sufficient. (Show this directly by checking when $f(\mathbf{x}|\theta)/f(\mathbf{y}|\theta)$ is free of θ. A different method will be given later.) Now note that $S(X_1, \ldots, X_n) = \sum_{i=1}^n (X_i - \bar{X})^2$ is ancillary. Either because, $S^2 \sim \chi^2_{n-1}$ which is free of θ, or because $X_i - \bar{X} = (X_i - \theta) - (\bar{X} - \theta) = Z_i - \bar{Z}$ where $Z_i = X_i - \theta$. Since θ is location parameter for X_i, distribution of Z_i is free of θ. Similarly, if $X_1, \ldots, X_n$ are i.i.d. $N(0, \sigma^2)$, then $V^2 = \sum_{i=1}^n X_i^2$ is sufficient and $T = \bar{X}/V$ is ancillary. Either, note that

$$nT^2 = \frac{n\bar{X}^2}{n\bar{X}^2 + \sum_{i=1}^n (X_i - \bar{X})^2} \sim \text{Beta}(\frac{1}{2}, \frac{n-1}{2}),$$

which is free of σ, or that

$$T = \frac{\bar{X}}{V} = \frac{\bar{X}/\sigma}{V/\sigma} = \frac{\bar{Z}}{V_Z},$$

where $Z_i = X_i/\sigma$ and $V_Z^2 = \sum_{i=1}^n Z_i^2$; Z_i is free of σ since it is a scale parameter of X_i.

In fact, here is a general result. Let $X_1, \ldots, X_n$ be i.i.d from a location-scale distribution with location μ and scale σ. Then, for any four integers a, b, c, and d (between 1 and n), the ratio

$$\frac{X_{(a)} - X_{(b)}}{X_{(c)} - X_{(d)}} = \frac{Z_{(a)} - Z_{(b)}}{Z_{(c)} - Z_{(d)}}$$

is ancillary because the right-hand side is expressed in terms of order statistics of Z_i's where $Z_i = (X_i - \mu)/\sigma$, $i = 1, \ldots, n$ are i.i.d. with a distribution free of μ and σ.

Example. Let $X_1, \ldots, X_n$ be i.i.d $U(\theta, \theta+1)$, $-\infty < \theta < \infty$. Then

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n|\theta) = \begin{cases} 1 & \text{if } \theta < x_{(1)} < \cdots < x_{(n)} < \theta + 1; \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} 1 & \text{if } x_{(n)} - 1 < \theta < x_{(1)}; \\ 0 & \text{otherwise.} \end{cases},$$

2

implying that $(X_{(1)}, X_{(n)})$ is sufficient for θ. For two sample points $\mathbf{x}$ and $\mathbf{y}$ (they must satisfy $x_{(1)} < x_{(n)} < x_{(1)} + 1$ and similar property for $\mathbf{y}$), consider the ratio $f(\mathbf{x}|\theta)/f(\mathbf{y}|\theta)$. This is a constant equal to 1 if $x_{(1)} = y_{(1)}$ and $x_{(n)} = y_{(n)}$. If these equalities do not hold, then there will exist θ for which $f(\mathbf{x}|\theta) > 0$ and $f(\mathbf{y}|\theta) = 0$ and some other θ for which $f(\mathbf{x}|\theta) = 0$ and $f(\mathbf{y}|\theta) > 0$. Then the ratio above will not be a constant function of θ. Therefore, $(X_{(1)}, X_{(n)})$ is minimal sufficient for θ. Then $((X_{(1)}+X_{(n)})/2, X_{(n)}-X_{(1)})$ which is a one-one function is also minimal sufficient. (Note they are equivalent statistics and provide the same partition of the sample space.) Now note that $R = X_{(n)} - X_{(1)} = (X_{(n)} - \theta) - (X_{(1)} - \theta) = Z_{(n)} - Z_{(1)}$, where $Z_i = X_i - \theta \sim U(0,1)$. Thus we see that R is ancillary even though it is part of the minimal sufficient statistics. Note from the following that $R \sim \text{Beta}(n-1, 2)$, which shows once again that it is free of θ.

$$
\begin{aligned}
P(X_{(1)} > x_{(1)}, X_{(n)} \le x_{(n)}) &= P(x_{(1)} < X_i \le x_{(n)} \;\forall\; i) \\
&= (x_{(n)} - x_{(1)})^n \text{ if } \theta < x_{(1)} < x_{(n)} < \theta + 1; \quad \text{so} \\
F_{X_{(1)},X_{(n)}}(x_{(1)}, x_{(n)}) &= P(X_{(1)} \le x_{(1)}, X_{(n)} \le x_{(n)}) \\
&= P(X_{(n)} \le x_{(n)}) - P(X_{(1)} > x_{(1)}, X_{(n)} \le x_{(n)}) \\
&= g(x_{(n)}) - (x_{(n)} - x_{(1)})^n \text{ if } \theta < x_{(1)} < x_{(n)} < \theta + 1.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
f_{X_{(1)},X_{(n)}}(x_{(1)}, x_{(n)}) &= \frac{\partial^2}{\partial x_{(1)} \partial x_{(n)}} F_{X_{(1)},X_{(n)}}(x_{(1)}, x_{(n)}) \\
&= n(n-1)(x_{(n)} - x_{(1)})^{n-2} \text{ if } \theta < x_{(1)} < x_{(n)} < \theta + 1.
\end{aligned}
$$

Taking $R = X_{(n)} - X_{(1)}$, $M = (X_{(1)} + X_{(n)})/2$, we get $X_{(1)} = (2M - R)/2$ and $X_{(n)} = (2M + R)/2$, with the Jacobian of the transformation equal to 1 (since $\begin{vmatrix} -\frac{1}{2} & 1 \\ \frac{1}{2} & 1 \end{vmatrix} = -1$), and hence

$$
f_{R,M}(r, m) = \begin{cases} n(n-1)r^{n-2} & \text{if } 0 < r < 1, \theta + \frac{r}{2} < m < \theta + 1 - \frac{r}{2}; \\ 0 & \text{otherwise, and} \end{cases}
$$

$$
f_R(r) = \int_{\theta+\frac{r}{2}}^{\theta+1-\frac{r}{2}} n(n-1)r^{n-2}\, dm = n(n-1)r^{n-2}(1 - r), 0 < r < 1.
$$

Let us state this as a general result.

Result. Let $X_1, \ldots, X_n$ be i.i.d from a location parameter family with cdf $F_X(x|\theta) = F_0(x - \theta)$, $-\infty < \theta < \infty$. Then $R = X_{(n)} - X_{(1)}$ is ancillary.

Proof. Let $Z_i = X_i - \theta$. Then Z_i has location 0 and cdf $F_Z(z) = F_0(z)$. Further,

$$
\begin{aligned}
F_R(r|\theta) &= P_\theta(R \leq r) = P(X_{(n)} - X_{(1)} \leq r) \\
&= P((Z_{(n)} + \theta) - (Z_{(1)} + \theta) \leq r) = P(Z_{(n)} - Z_{(1)} \leq r),
\end{aligned}
$$

which is free of θ.

Result. Let $X_1, \ldots, X_n$ be i.i.d from a scale parameter family with cdf $F_X(x|\sigma) = F_1(x/\sigma)$, $\sigma > 0$. Then any statistic, $h\left(\frac{X_1}{X_n}, \cdots, \frac{X_{n-1}}{X_n}\right)$ is ancillary.

Proof. Let $Z_i = X_i - \sigma$. Then Z_i has scale 1 and cdf $F_Z(z) = F_1(z)$. Note that

$$
\begin{aligned}
h\left(\frac{X_1}{X_n}, \cdots, \frac{X_{n-1}}{X_n}\right) &= h\left(\frac{X_1/\sigma}{X_n/\sigma}, \cdots, \frac{X_{n-1}/\sigma}{X_n/\sigma}\right) \\
&= h\left(\frac{Z_1}{Z_n}, \cdots, \frac{Z_{n-1}}{Z_n}\right),
\end{aligned}
$$

which is free of σ.