

Theorem. Let $P_\theta, \theta \in \Theta$ be a k -parameter exponential family with density $f(\mathbf{x}|\theta) = \exp\left(\sum_{j=1}^k c_j(\theta)T_j(\mathbf{x}) + d(\theta) + S(\mathbf{x})\right) I_A(\mathbf{x})$. Suppose $\{\mathbf{c}(\theta) = (c_1(\theta), \dots, c_k(\theta)), \theta \in \Theta\}$ contains an open set (open rectangle) in \mathcal{R}^k . (This property is called full-rank.) Then $\mathbf{T}(\mathbf{X}) = (T_1, \dots, T_k)$ is complete sufficient.

Proof. See Lehmann: Testing Statistical Hypotheses for a proof involving uniqueness of Laplace transforms. We will use the result.

Example. Let X_1, \dots, X_n be i.i.d $N(\mu, \sigma^2)$, $-\infty < \mu < \infty$, $\sigma^2 > 0$. $T(\mathbf{X}) = (\bar{X}, \sum_{i=1}^n (X_i - \bar{X})^2)$ is sufficient. Since the set of $c(\theta) = (\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2})$ for $-\infty < \mu < \infty$, $\sigma^2 > 0$ is $\mathcal{R}^1 \times \mathcal{R}^+$, an open set, $T(\mathbf{X})$ is complete also.

Therefore, since $E(\bar{X}) = \mu$ and $\bar{X} = h(T(\mathbf{X}))$, we obtain that \bar{X} is the unique UMVUE of μ .

Further, $\sum_{i=1}^n (X_i - \bar{X})^2 \sim \sigma^2 \chi_{n-1}^2$, and hence

$$E\left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right] = \sigma^2, \forall (\mu, \sigma^2).$$

Therefore, $\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$, which is a function of $T(\mathbf{X})$ alone, is the unique UMVUE of σ^2 .

Suppose we want to estimate $q(\mu, \sigma^2) = \mu/\sigma$. Then, for $n \geq 3$, the UMVUE of $q(\mu, \sigma^2)$ is

$$c(n) \frac{\bar{X}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}},$$

where $c(n)$ may be found from

$$E\left[\frac{\bar{X}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}\right] = E(\bar{X})E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right]^{-1/2},$$

where $E(\bar{X}) = \mu$ and $\sum_{i=1}^n (X_i - \bar{X})^2/\sigma^2 \sim \Gamma(1/2, (n-1)/2)$.

Example. Let X_1, \dots, X_n be i.i.d $N(\theta, \theta^2)$, $\theta > 0$. Then $T(\mathbf{X}) = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ is (minimal) sufficient for θ , but it is not complete. Therefore it is not possible to apply L-S to find the UMVUE of θ .

There is an interesting technical theorem, due to D. Basu, which establishes independence of a sufficient statistic and an ancillary statistic. The result is useful in many calculations. Recall that parts of a minimal sufficient statistic may be ancillary, so conditions are needed for this to happen.

Theorem (Basu). Suppose T is complete sufficient for $\{P_\theta, \theta \in \Theta\}$. Let S be any ancillary statistic. Then T and S are independent for all θ .

Proof. Because T is sufficient, the conditional probability of S being in some set B given T is free of θ and may be written as $P_\theta(S \in B|T) = \phi(T)$. Since S is ancillary, $E_\theta(\phi(T)) = P_\theta(S \in B) = c$, where c is a constant. Consider a B for which $0 < c < 1$. Let $\psi(T) = \phi(T) - c$. Then $E_\theta\psi(T) = 0$ for all θ , implying $\psi(T) = 0$ (with probability one), i.e., $P_\theta(S \in B|T) = P_\theta(S \in B)$.

Alternatively, let $f(t, s|\theta)$ be the joint density of (T, S) and $F_S(s)$ be the cdf of S . Then F_S is free of θ since S is ancillary. Let $f_T(t|\theta)$ be the density of T and $F_{S|T}(s)$ be the conditional cdf of $S|T$ (which is again free of θ since T is sufficient). Note that, for any s ,

$$\begin{aligned} \int F_S(s) f_T(t|\theta) dt &= F_S(s) \int f_T(t|\theta) dt = F_S(s). \\ \int F_{S|T=t}(s) f_T(t|\theta) dt &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^s f_{S|T=t}(u) du \right] f_T(t|\theta) dt \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^s f_{S|T=t}(u) f_T(t|\theta) du dt \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^s f_{S,T}(u, t) du dt \\ &= \int_{-\infty}^s \left(\int_{-\infty}^{\infty} f_{S,T}(u, t) dt \right) du \\ &= \int_{-\infty}^s f_S(u) du = F_S(s). \end{aligned}$$

Therefore,

$$\int_{-\infty}^{\infty} [F_S(s) - F_{S|T=t}(s)] f_T(t|\theta) dt = 0, \forall \theta.$$

Fix s and let $h(T) = F_S(s) - F_{S|T=t}(s)$, which involves T but is totally free of θ . Then $E_\theta h(T) = 0$ for all θ . Since T is complete, $h \equiv 0$. That means, $F_{S|T} \equiv F_S$, implying independence of S and T for all θ .

Example. Suppose X_1, X_2, \dots, X_n are i.i.d. $N(\mu, \sigma^2)$. Then \bar{X} and $S^2 = \sum (X_i - \bar{X})^2$ are independent. To prove this, treat σ^2 as fixed to start with and μ as the parameter. Then \bar{X} is complete sufficient and $S^2 = \sum (X_i - \bar{X})^2 = \sum [(X_i - \mu) - (\bar{X} - \mu)]^2 = \sum (Z_i - \bar{Z})^2$ is ancillary. Hence \bar{X} and S^2 are independent for each σ^2 by Basu's theorem.

Example. Suppose X_1, X_2, \dots, X_n are i.i.d $U(\theta_1, \theta_2)$. Then for any $1 < r < n$, $Y = (X_{(r)} - X_{(1)}) / (X_{(n)} - X_{(1)})$ is independent of $(X_{(1)}, X_{(n)})$. This follows

because Y is ancillary.

It is shown below that a complete sufficient statistic is minimal sufficient. In general, the converse isn't true. Also, technically, neither may exist or only one of them may exist.

Theorem. A (bdd) complete sufficient statistic is minimal sufficient, assuming minimal sufficient statistic exists.

Proof. Let T be minimal sufficient and U be complete sufficient. Then $T = h(U)$ for some function h since minimal sufficiency provides coarser partition. We need to show that T and U are equivalent statistics (i.e., produce the same partition). It is enough to show that for all (integrable) ψ ,

$$E[\psi(U)|T] = \psi(U).$$

(Note that $T = h(U)$ and the above requirement is simply that averaging $\psi(U)$ where $h(U)$ is fixed, reproduces $\psi(U)$.)

Suppose not. That is, let there exist ψ such that $E[\psi(U)|h(U)]$ is not identical to $\psi(U)$. Define

$$k(U) = \psi(U) - E[\psi(U)|h(U)].$$

Then $k \neq 0$ but

$$\begin{aligned} E[k(U)] &= E\{\psi(U) - E[\psi(U)|h(U)]\} \\ &= E(\psi(U)) - E\{E[\psi(U)|h(U)]\} \\ &= E(\psi(U)) - E\{\psi(U)\} = 0. \end{aligned}$$

However, U is complete!

Corollary. Let $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ be a k -parameter exponential family with density

$$f(\mathbf{x}|\theta) = \exp\left(\sum_{i=1}^k c_i(\theta)T_i(\mathbf{x} + d(\theta) + S(\mathbf{x}))\right) I_A(\mathbf{x}),$$

where $C = \{(c_1(\theta), \dots, c_k(\theta)) : \theta \in \Theta\}$ contains an open set. Then (T_1, \dots, T_k) is minimal sufficient.

Proof. (T_1, \dots, T_k) is complete sufficient as remarked previously, hence minimal sufficient.