Lecture 6: Exponential Family, MLE, and Sufficiency

6.1 Exponential family

Definition 6.1 The probability density function of an exponential family is given by

$$\text{pdf}(x;\eta) = \exp\{\eta T(x) - A(\eta) + S(x)\}, \quad x \in \mathcal{X}, \eta \in \Omega$$

satisfying the following properties:

(i) the support $\mathcal{X} = \{x : \text{pdf}(x;\eta) > 0\}$ does not change with respect to the parameter η.
(ii) the parameter space Ω is an open interval on the real line.
(iii) $\text{Var}(T(X_1)) > 0$, i.e., $T(x)$ is not a constant function.

where η is the "natural" parameter, and $A(\eta)$ is the "log-partition" or "log-normalization" function.

Remark 6.1 An exponential family is an important class of probability distributions with a specific form that exhibits mathematical conveniences (e.g., straightforward calculation of expectations and covariances through differentiation). More importantly, we will show that this class of distributions naturally determines complete and sufficient statistics in the later sections. In this section, even though we only consider continuous variable cases for the proof, it can include discrete variable cases.

Theorem 6.1 *If a random sample $X_1, X_2, \cdots, X_n \sim f_\eta$, and f belongs to an exponential family. Then, the likelihood equation is given by*

$$\dot{A}(\eta) = \frac{\partial A(\eta)}{\partial \eta} = \mathbb{E}_\eta[T(X_1)] = \frac{1}{n}\sum_{i=1}^{n} T(x_i), \quad \eta \in \Omega. \tag{1}$$

If the solution of (1) $\hat{\eta} = \hat{\eta}(x_1, \ldots, x_n)$ exists, then $\hat{\eta}$ is the maximum likelihood estimate of η. (Note: $\ddot{A} = \frac{\partial^2}{\partial \eta^2} A(\eta)$)

Proof. Omitted □

Remark 6.2 In many cases, the parameter η used in the expression of the exponential family is given by a one-to-one transformation g as $\eta = g(\theta)$ for some $\theta \in \Theta$. The pdf parameterized with the natural parameter θ is given by

$$\text{pdf}(x;\theta) = \exp\left\{g(\theta)T(x) - A(g(\theta)) + S(x)\right\}, \quad x \in \mathcal{X}, \quad \theta \in \Theta.$$

Therefore, if the distribution belongs to the exponential family, the maximum likelihood estimator of θ is obtained using the likelihood equation:

$$\mathbb{E}_\theta[T(X_1)] = \frac{1}{n}\sum_{i=1}^{n} T(x_i), \quad \theta \in \Theta,$$

from Theorem 5.3.

Example 6.1 Find the MLE for the parameters in each of the following using Theorem 6.1:

(a) $X_1, \ldots, X_n \overset{iid}{\sim} \text{Ber}(p)$, with $0 < p < 1$

Solution. The pdf is given by

$$\text{pdf}(x; p) = p^x(1-p)^{1-x} = \exp\left\{\left(\log\frac{p}{1-p}\right)x + \log(1-p)\right\}, \quad x = 0, 1.$$

We apply the transformation $\eta = \log\frac{p}{1-p}$. Note that in order to continue, we need $\log(1-p)$ in terms of η. We know $e^\eta = \frac{p}{1-p}$. Solving, we get $p = \frac{e^\eta}{1+e^\eta}$. Plugging in, the pdf with η becomes

$$\text{pdf}(x; \eta) = \exp\left\{\eta x - \log\left(1 + e^\eta\right)\right\}, \quad x = 0, 1, \quad -\infty < \eta < \infty.$$

Checking the conditions, we see that the pdf is an exponential family, therefore we apply Theorem 6.1 and Remark 6.2

$$\frac{\partial A(\eta)}{\partial \eta} = \frac{e^\eta}{1+e^\eta} = p = \mathbb{E}_p[X_1] = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Therefore, the MLE is $\hat{p}_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n} X_i$. ■

(b) $X_1, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$, with $\lambda > 0$

Solution. The pdf is given by

$$\text{pdf}(x; \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \ldots$$

When we apply the transformation $\eta = \log\lambda$, the probability density function becomes

$$\text{pdf}(x; \eta) = \exp\left(\eta x - e^\eta - \log x!\right), \quad x = 0, 1, \ldots, \quad -\infty < \eta < \infty.$$

For a random sample of size n, the likelihood equation using the parameter λ is given by

$$\frac{1}{n}\sum_{i=1}^{n} x_i = E_\lambda(X_1) = \lambda, \quad \lambda > 0.$$

Therefore, when $\sum_{i=1}^{n} x_i > 0$, the maximum likelihood estimator (MLE) of λ is

$$\hat{\lambda}_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

■

(c) $X_1, \ldots, X_n \overset{iid}{\sim} \text{Geo}(p)$ with $0 < p < 1$

Solution. The pdf is given by

$$\text{pdf}(x; p) = (1-p)^{x-1}p, \quad x = 1, 2, \ldots$$

When we apply the transformation $\eta = \log(1-p)$, the probability density function becomes

$$\text{pdf}(x; \eta) = \exp\left(\eta x - \eta + \log(1 - e^{\eta})\right), \quad x = 0, 1, \ldots, \quad -\infty < \eta < 0$$

For a random sample of size n, the likelihood equation using the parameter p is given by

$$\frac{1}{n}\sum_{i=1}^{n} x_i = E_p(X_1) = \frac{1}{p}, \quad 0 < p < 1.$$

Therefore, when $\sum_{i=1}^{n} x_i/n > 1$, the maximum likelihood estimator (MLE) of p is

$$\hat{p}_{\text{MLE}} = \frac{1}{\sum_{i=1}^{n} x_i/n}$$

■

(d) $X_1, \ldots, X_n \overset{iid}{\sim} \text{Exp}(\theta)$, with $\theta > 0$.

Solution. The pdf is given by

$$\text{pdf}(x; \theta) = \frac{1}{\theta}e^{-x/\theta}, \quad x > 0.$$

When we apply the transformation $\eta = -\frac{1}{\theta}$, the probability density function becomes

$$\text{pdf}(x; \eta) = \exp\left(\eta x + \log(-\eta)\right), \quad x > 0, \quad \eta < 0$$

For a random sample of size n, the likelihood equation using the parameter θ is given by

$$\frac{1}{n}\sum_{i=1}^{n} x_i = E_\theta(X_1) = \theta, \quad \theta > 0.$$

Therefore, the maximum likelihood estimator (MLE) of θ is

$$\hat{\theta}_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

■

(e) $X_1, \ldots, X_n \overset{iid}{\sim} \text{Pareto}(1, \theta)$, with $\theta > 0$

Solution. The pdf is given by

$$\text{pdf}(x; \theta) = \theta x^{-\theta-1}, \quad x > 1.$$

In this case, $\eta = \theta$, so no transformation of the parameter is required. The probability density function becomes

$$\text{pdf}(x; \theta) = \exp\left(\theta(-\log x) + \log\theta - \log x\right), \quad x > 1, \quad \theta > 0.$$

For a random sample of size n, the likelihood equation using the parameter θ is given by

$$\frac{1}{n}\sum_{i=1}^{n}(-\log x_i) = \dot{A}(\hat{\theta}) = -\frac{1}{\hat{\theta}}, \quad \hat{\theta} > 0.$$

At the same time, we have

$$X_1^{-\theta} \sim \text{U}(0,1), \quad \theta \log X_1 \sim \text{Exp}(1), \quad E_\theta(-\log X_1) = -\frac{1}{\theta},$$

thus,

$$\dot{A}(\theta) = \mathbb{E}_\theta(-\log X_1).$$

We can conclude that

$$\hat{\theta}_{\text{MLE}} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\log x_i}.$$

■

Definition 6.2 The pdf of the multivariate exponential family is given by

$$\text{pdf}(x; \boldsymbol{\eta}) = \exp\left(\sum_{j=1}^{k}\eta_j T_j(x) - A(\boldsymbol{\eta}) + S(x)\right), \quad x \in \mathcal{X}, \quad \boldsymbol{\eta} = (\eta_1, \ldots, \eta_k)^T \in \Omega$$

with the following properties:

(i) the support of the distribution $\mathcal{X} = \{x : \text{pdf}(x; \eta) > 0\}$ does not change with respect to η.

(ii) the parameter space Ω contains an open k-dimensional region in $\mathbb{R}^k$ defined by the Cartesian product of intervals $(a_1, b_1) \times \cdots \times (a_k, b_k)$.

(iii) For any real constants $\boldsymbol{c} = (c_1, \ldots, c_k)^T$, the linear combination $\boldsymbol{c}^T\boldsymbol{T}(x) = c_1T_1(x) + \cdots + c_kT_k(x)$ is not a constant. That is, for any nonzero vector $\boldsymbol{c} \in \mathbb{R}^k$, the variance is positive:

$$\mathrm{Var}(\boldsymbol{c}^T\boldsymbol{T}(X)) > 0 \quad \forall\, \boldsymbol{c} \in \mathbb{R}^k,\, \boldsymbol{c} \neq \boldsymbol{0}.$$

Theorem 6.1 then applies. See the next example for how.

Example 6.2 The pdf for $X \sim N(\mu, \sigma^2)$ is

$$\begin{aligned}
\mathrm{pdf}(x;\boldsymbol{\theta}) &= (2\pi\sigma^2)^{-1/2}\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right), \quad -\infty < x < \infty, \quad \boldsymbol{\theta} = (\mu, \sigma^2)^T \\
&= \exp\left\{-\frac{1}{2}\log\big(2\pi\sigma^2\big) - \frac{1}{2\sigma^2}(x^2 - 2\mu x + \mu^2)\right\} \\
&= \exp\left\{\frac{\mu}{\sigma^2}\cdot x + \frac{-1}{2\sigma^2}\cdot x^2 - \left[\frac{1}{2}\log\big(\sigma^2\big) + \frac{\mu^2}{2\sigma^2}\right] - \frac{1}{2}\log(2\pi)\right\}
\end{aligned}$$

We can define the following transformations

$$\eta_1 = \frac{\mu}{\sigma^2}, \quad \eta_2 = \frac{-1}{2\sigma^2},$$

where the pdf is now of the form

$$\mathrm{pdf}(x;\eta_1,\eta_2) = \exp\left(\eta_1 x + \eta_2 x^2 - A(\eta_1,\eta_2) + S(x)\right), \quad -\infty < x < +\infty, \quad -\infty < \eta_1 < +\infty, \quad \eta_2 < 0.$$

We now need $A(\eta_1, \eta_2)$ by writing $\frac{1}{2}\log\big(\sigma^2\big) + \frac{\mu^2}{2\sigma^2}$ in terms of η_1 and η_2. With some algebra,

$$A(\eta_1,\eta_2) = -\frac{1}{2}\log(-2\eta_2) - \frac{\eta_1^2}{4\eta_2}.$$

Let $\boldsymbol{\eta} = (\eta_1, \eta_2)$. The log-likelihood function for a sample of size n is given by

$$l(\boldsymbol{\eta}) = \eta_1\sum_{i=1}^n x_i + \eta_2\sum_{i=1}^n x_i^2 - nA(\boldsymbol{\eta}),$$

and the maximum likelihood estimation equations using the parameters $\theta = (\mu, \sigma^2)^T$ are

$$E_\theta(X_1) = \frac{1}{n}\sum_{i=1}^n x_i, \quad E_\theta(X_1^2) = \frac{1}{n}\sum_{i=1}^n x_i^2, \quad \theta = (\mu,\sigma^2)^T, \quad -\infty < \mu < \infty, \quad \sigma^2 > 0.$$

Therefore, the likelihood equation regarding μ and σ^2 are given by

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i, \quad \hat{\sigma}^2 + (\hat{\mu})^2 = \frac{1}{n}\sum_{i=1}^n x_i^2,$$

which implies $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i, \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ are the MLEs.

6.2 Sufficient statistics

Remark 6.3 When estimating parameters using a random sample $X_1, \ldots, X_n$, if the estimator can be fully and precisely determined by using only some function of the sample, it may be advantageous because we only need the value of the function.

This concept of data reduction leads to the notion of sufficiency.

Example 6.3 Can the observed data from independent Bernoulli trials X_1, X_2 with parameter θ (where $0 < \theta < 1$) be fully reconstructed from $Y = X_1 + X_2$ to estimate θ? Since X_1 and X_2 are Bernoulli trials, $Y = X_1 + X_2$ can take values in $\{0, 1, 2\}$:

(i) If $Y = 0$, then $X_1 = 0$ and $X_2 = 0$.

(ii) If $Y = 1$, the possibilities are $(X_1, X_2) = (1, 0)$ or $(0, 1)$.

(iii) If $Y = 2$, then $X_1 = 1$ and $X_2 = 1$.

The probabilities are given by:

(i) $P(X_1 = 0, X_2 = 0 \mid Y = 0) = 1$

(ii) $P(X_1 = 1, X_2 = 0 \mid Y = 1) = P(X_1 = 0, X_2 = 1 \mid Y = 1) = \frac{1}{2}$

(iii) $P(X_1 = 1, X_2 = 1 \mid Y = 2) = 1$

Since these probabilities do not depend on θ, the sum $Y = X_1 + X_2$ contains all the information about θ, making it a sufficient statistic. In words, knowing Y is sufficient for θ regardless of the order of success or fail.

Definition 6.3 For a probability density function $f(x; \theta)$ with parameter $\theta \in \Omega$ from the sample $X_1, \ldots, X_n$, the statistic $Y = u(X_1, \ldots, X_n)$ is **sufficient** for θ if

$$P_{\theta_1}((X_1, \ldots, X_n)^T \in A \mid Y = y) = P_{\theta_2}((X_1, \ldots, X_n)^T \in A \mid Y = y) \quad \forall A, y, \text{ and } \forall \theta_1, \theta_2 \in \Omega.$$

Equivalently, if the conditional distribution of $(X_1, \ldots, X_n)^T$ given the statistic $Y = u(X_1, \ldots, X_n) = y$ does not depend on $\theta \in \Omega$, we call Y a **sufficient** statistic for $\theta \in \Omega$.

Example 6.4 Let $X_1, \ldots, X_n \sim \text{Ber}(\theta)$, where $0 < \theta < 1$. Define $Y := X_1 + \cdots + X_n$. We know $Y \sim \text{Binomial}(n, \theta)$. Thus, given $Y = X_1 + \cdots + X_n = y$, the conditional probability of the random sample $(X_1, \ldots, X_n)^T$ is

$$\begin{aligned}
P_\theta(X_1 = x_1, \ldots, X_n = x_n | X_1 + \cdots + X_n = y) &= \frac{P_\theta(X_1 + \cdots + X_n = y \mid X_1 = x_1, \ldots, X_n = x_n) \cdot P_\theta(X_1 = x_1, \ldots, X_n = x_n)}{P_\theta(X_1 + \cdots + X_n = y)} \\
&= \frac{\prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}}{\binom{n}{y}\theta^y(1-\theta)^{n-y}} \cdot \mathbf{1}\{x_1 + \cdots + x_n = y\} \\
&= \frac{1}{\binom{n}{y}} \cdot \mathbf{1}\{x_1 + \cdots + x_n = y\}, \quad \forall \theta \in (0,1).
\end{aligned}$$

Hence, $Y = X_1 + \cdots + X_n$ is a sufficient statistic for $\theta \in (0,1)$, according to the definition. This begs the question, given just the likelihood function, how can we find a sufficient statistic?

Theorem 6.2 ***(Factorization Theorem)*** *Let $f(x;\theta)$ be the probability density function for a random sample $X_1, \ldots, X_n$ with parameter $\theta \in \Omega$. Then $Y = u(X_1, \ldots, X_n)$ is a sufficient statistic for $\theta \in \Omega$ if and only if there exist functions k_1, k_2 satisfying*

$$\prod_{i=1}^n f(x_i;\theta) = k_1(u(\boldsymbol{x}),\theta)k_2(\boldsymbol{x}), \quad \forall \boldsymbol{x} = (x_1, \ldots, x_n)^T, \quad \forall \theta \in \Omega.$$

Proof. This is a general result, but we will prove it for a discrete case.
($\Leftarrow$) Suppose that there exist such functions, we have

$$P_\theta(\boldsymbol{X} = \boldsymbol{x}) = \prod_{i=1}^n f(x_i;\theta) = k_1(u(\boldsymbol{x}),\theta)k_2(\boldsymbol{x}), \quad \forall x = (x_1, \ldots, x_n)^T, \quad \forall \theta \in \Omega.$$

Thus, for all $\theta \in \Omega$ and $y = u(x_1, \ldots, x_n)$, the following holds:

$$\begin{aligned}
P_\theta(\boldsymbol{X} = \boldsymbol{x} | Y = y) &= \frac{P_\theta(\boldsymbol{X} = \boldsymbol{x}, Y = u(\boldsymbol{X}) = y)}{P_\theta(Y = u(\boldsymbol{X}) = y)} \\
&= \frac{P_\theta(\boldsymbol{X} = \boldsymbol{x}) \cdot \mathbf{1}\{u(\boldsymbol{x}) = y\}}{\sum_{\boldsymbol{z}:u(\boldsymbol{z})=y} P_\theta(\boldsymbol{X} = \boldsymbol{z})} \\
&= \frac{k_1(u(\boldsymbol{x}),\theta)k_2(\boldsymbol{x}) \cdot \mathbf{1}\{u(\boldsymbol{x}) = y\}}{\sum_{\boldsymbol{z}:u(\boldsymbol{z})=y} k_1(u(\boldsymbol{z}),\theta)k_2(\boldsymbol{z})} \\
&= \frac{k_1(y,\theta)k_2(\boldsymbol{x}) \cdot \mathbf{1}\{u(\boldsymbol{x}) = y\}}{k_1(y,\theta)\sum_{\boldsymbol{z}:u(\boldsymbol{z})=y} k_2(\boldsymbol{z})} \\
&= \frac{k_2(\boldsymbol{x}) \cdot \mathbf{1}\{u(\boldsymbol{x}) = y\}}{\sum_{\boldsymbol{z}:u(\boldsymbol{z})=y} k_2(\boldsymbol{z})}.
\end{aligned}$$

This conditional probability no longer depends on θ, thus proving that $Y = u(X_1, \ldots, X_n)$ is a sufficient statistic for $\theta \in \Omega$.
($\Rightarrow$) If $Y = u(\boldsymbol{X}) = u(X_1, \ldots, X_n)$ is a sufficient statistic for $\theta \in \Omega$, then the conditional probability $P_\theta(\boldsymbol{X} = \boldsymbol{x} \mid Y = y)$ must not depend on $\theta \in \Omega$ for all $\boldsymbol{x}$ and y. Therefore, we have

$$P_\theta(\boldsymbol{X} = \boldsymbol{x} \mid Y = u(\boldsymbol{X}) = y) = k_2(\boldsymbol{x})$$

Now, let

$$P_\theta(Y = u(\boldsymbol{x})) = k_1(u(\boldsymbol{x}), \theta)$$

Thus, for all $\theta \in \Omega$ and for $\boldsymbol{x} = (x_1, \ldots, x_n)^T$, the following holds

$$P_\theta(\boldsymbol{X} = \boldsymbol{x}) = P_\theta(\boldsymbol{X} = \boldsymbol{x}, Y = u(\boldsymbol{X}) = u(\boldsymbol{x})) = P(\boldsymbol{X} = \boldsymbol{x} \mid Y = u(\boldsymbol{x}))P_\theta(Y = u(\boldsymbol{x})) = k_2(\boldsymbol{x})k_1(u(\boldsymbol{x}), \theta).$$

Therefore, we can write

$$\prod_{i=1}^{n} f(x_i; \theta) = P_\theta(\boldsymbol{X} = \boldsymbol{x}) = P(\boldsymbol{X} = \boldsymbol{x} \mid Y = u(\boldsymbol{x}))P_\theta(Y = u(\boldsymbol{x})) = k_1(u(\boldsymbol{x}), \theta)k_2(\boldsymbol{x}),$$

which completes the proof. □

Example 6.5 Suppose $X_1, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\theta)$, the joint pmf is

$$\prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} \left(\frac{e^{-\theta}\theta^{x_i}}{x_i!} \right) = \prod_{i=1}^{n} e^{-\theta} \frac{\theta^{x_i}}{x_i!} = e^{-n\theta}\theta^{\sum_{i=1}^n x_i} \prod_{i=1}^{n} \frac{1}{x_i!}.$$

Then, define

$$e^{-n\theta}\theta^{\sum_{i=1}^n x_i} = k_1 \left(\sum_{i=1}^{n} x_i, \theta \right),$$

which leads to

$$\prod_{i=1}^{n} f(x_i; \theta) = k_1 \left(\sum_{i=1}^{n} x_i, \theta \right) \prod_{i=1}^{n} \frac{1}{x_i!}.$$

Therefore, $u(X) = \sum_{i=1}^n X_i$ is a sufficient statistic for $\theta > 0$.

Theorem 6.3 *Suppose $X_1, X_2, \ldots, X_n$ follows a distribution which belongs to a multivariate exponential family, whose probability distribution is given by*

$$f(x; \theta) = \exp \left(\sum_{j=1}^{k} g_j(\theta) T_j(x) - A(g(\theta)) + S(x) \right), \quad x \in \mathcal{X}, \theta \in \Omega.$$

Then, we have a sufficient statistic for $\theta \in \Omega$ given by

$$\sum_{i=1}^{n} \boldsymbol{T}(X_i) = \left(\sum_{i=1}^{n} T_1(X_i), \ldots, \sum_{i=1}^{n} T_k(X_i) \right)^T.$$

Proof. The joint probability density function of the random samples $X_1, X_2, \ldots, X_n$ is given by

$$\prod_{i=1}^{n} f(x_i;\theta) = \exp\left(\sum_{j=1}^{k} g_j(\theta) \sum_{i=1}^{n} T_j(x_i) - nA(g(\theta))\right) \exp\left(\sum_{i=1}^{n} S(x_i)\right)$$

Since the likelihood function can be factorized as

$$\prod_{i=1}^{n} f(x_i;\theta) = \exp\left(\sum_{j=1}^{k} g_j(\theta) \sum_{i=1}^{n} T_j(x_i) - nA(g(\theta))\right) \cdot h(x)$$

where $h(x) = \exp\left(\sum_{i=1}^{n} S(x_i)\right)$, we conclude that

$$\sum_{i=1}^{n} \boldsymbol{T}(X_i) = \left(\sum_{i=1}^{n} T_1(X_i), \ldots, \sum_{i=1}^{n} T_k(X_i)\right)^T$$

is a sufficient statistic for $\theta \in \Omega$, by the factorization theorem. □

Remark 6.4 Note that sufficient statistics are not unique. Any one-to-one function of a sufficient statistic is a sufficient statistic.

Theorem 6.4 *Let $f(x;\theta)$ be the probability density function for a random sample $X_1, \ldots, X_n$ from a population with parameter $\theta \in \Omega$. Suppose that $Y = u(X_1, \ldots, X_n)$ is a sufficient statistic for $\theta \in \Omega$. Then, any one-to-one function of Y, $W = g(Y)$, is also a sufficient statistic for $\theta \in \Omega$.*

Proof. Since $Y = u(X_1, \ldots, X_n)$ is a sufficient statistic for $\theta \in \Omega$, the conditional probability:

$$P_\theta\left((X_1, \ldots, X_n)^T \in A \mid Y = y\right)$$

does not depend on $\theta \in \Omega$ for all values of y. Now, let $W = g(Y)$ be a one-to-one function of Y, with $w = g(y)$ corresponding to the value w in the image of the function. Then, the conditional probability:

$$P_\theta\left((X_1, \ldots, X_n)^T \in A \mid W = w\right) = P_\theta\left((X_1, \ldots, X_n)^T \in A \mid Y = g^{-1}(w)\right).$$

Since Y is a sufficient statistic, this conditional probability does not depend on θ. Therefore, the statistic $W = g(Y)$ is also a sufficient statistic for $\theta \in \Omega$. □