Econ 240A: Sufficiency, Minimal Sufficiency and the Exponential Family of Distributions*

Chuan Goh

29 February 2001

This handout expands on the text by presenting a brief and hopefully useful discussion of the concepts of sufficiency, minimal sufficiency, and completeness, along with their relationship to the exponential family of distributions. Consider a random variable $X \sim \mathcal{P}_\theta$, where $\mathcal{P}_\theta$ refers to the class of probability distributions indexed by the parameter θ.

Definition 1 (Sufficiency). *A statistic $T(X)$ is sufficient for θ iff the conditional distribution of X given $T = t$ does not depend on θ.*

Deterimining the sufficiency of a statistic T via the definition can be cumbersome. Thankfully, there is a much simpler criterion.

Theorem 1 (Factorization Criterion). *T is sufficient for the parameter θ of the family of distributions generating X iff the densities of X satisfy*

$$p_\theta(x) = g_\theta[T(x)]h(x).$$

Proof. See Lehmann (1986), Section 2.6, Theorem 8 and Corollary 1. $\square$

Suppose we have an iid sequence of random variables $X_1, \ldots, X_n \sim \mathcal{P}_\theta$. Let $X = (X_1, \ldots, X_n)$ be the *data* for the experiment at hand. What information does a sufficient statistic $T(X)$ contain regarding the unknown parameter θ given a realization of the dataset $X = x$?

Suppose we record the realization of the statistic $T(X) = t$ but somehow misplace the original observations $X = x$. All is not lost as we can use a pseudo-random number generator to generate a dataset X' distributed according to the conditional distribution of X given t which by definition does not depend on θ.[1] It follows that the unconditional distribution of X' is the same as that of the original data X, ie., for any set A, $P_\theta(X \in A) = P_\theta(X' \in A)$. Therefore we can say that knowledge of T alone gives us the ability to replicate the random process that generated the data for the original experiment and so reconstruct misplaced data from the experiment. Alternatively, we can say that a sufficient

*This handout consists mainly of portions of Sections 1.5 and 1.6 of Lehmann and Casella (1998).

[1] Obviously, it would not be possible to do this if we needed to know the actual value of θ.

statistic allows one to compress the data without losing relevant information about the unknown parameter: if T is sufficient and there exist a function h and a statistic U such that $T = h(U)$, U cannot contain less information about θ than T and so U must also be sufficient. Moreover T provides a greater degree of data compression than U unless h is 1-to-1, in which case T and U are equivalent.

Example 1 (Normal data). *Suppose $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$. Their joint density is*

$$P_{\mu,\sigma^2}(x) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^{n} x_i - \frac{n}{2\sigma^2}\mu^2\right].$$

By the factorization criterion $T = (\sum X_i^2, \sum X_i)$ is sufficient for $\theta = (\mu, \sigma^2)$. Clearly T is equivalent to $T' = (\bar{X}, S^2)$, where $\bar{X} = n^{-1}\sum X_i$ and $S^2 = (n-1)^{-1}\sum(X_i - \bar{X})^2$.

Example 2 (Different sufficient statistics). *Let $X_1, \ldots, X_n$ be iid $N(0, \sigma^2)$. Then the statistics*

$$\begin{aligned}
T_1(X) &= (X_1, \ldots, X_n) \\
T_2(X) &= (X_1^2, \ldots, X_n^2) \\
T_3(X) &= (X_1^2 + \cdots + X_m^2, X_{m+1}^2 + \cdots + X_n^2) \\
T_4(X) &= X_1^2 + \cdots + X_n^2
\end{aligned}$$

are all sufficient for σ^2. T_i provides an increasing degree of data compression as i increases.

It is natural to ask if there is a statistic T that gives the maximal degree of data compression vis-à-vis θ. This leads to the notion of minimal sufficiency.

Definition 2 (Minimal Sufficiency). *A sufficient statistic T is minimal sufficient if for any sufficient statistic U there exists a function h such that $T = h(U)$.*

Construction of a minimal sufficient statistic is fairly straightforward. Consider the following lemma and theorem:

Lemma 1. *For any fixed θ and θ_0 a statistic U is sufficient iff $\frac{p_\theta(x)}{p_{\theta_0}(x)}$ is a function only of $U(x)$.*

Proof. Exercise. (This isn't too difficult.) $\square$

Theorem 2. *Consider a family of $s+1$ densities $p_i, i = 0, 1, 2, \ldots, s$ with a common support. Then $T(X) = \left(\frac{p_1(X)}{p_0(X)}, \cdots, \frac{p_s(X)}{p_0(X)}\right)$ is minimal sufficient.*

Proof. By the previous lemma we note that a statistic U is sufficient iff T is a function of U. But this is just the definition of minimal sufficiency applied to T. □

Although a minimal sufficient statistic provides in some sense an optimal degree of data compression it is still possible for it to contain much "extra" or *ancillary* material that does not provide by itself any information about θ.

Definition 3 (Ancillarity). *A statistic V is ancillary if its distribution does not depend on θ. V is first-order ancillary if the expectation $E_\theta[V(X)]$ does not depend on θ (i.e., $E_\theta[V(X)]$ is constant).*

A bit of thought will lead us to the idea that a sufficient statistic T that provides the most efficient degree of data compression will have the property that no nonconstant transformation of the statistic $h(T)$ is ancillary or first-order ancillary.[2] This notion is referred to as *completeness.*

Definition 4 (Completeness). *A statistic T is complete for θ if for any θ $E_\theta[h(T)] = c$ implies $h(T) = c$ with probability one. By subtracting c this condition is equivalent to*

$$E_\theta[h(T)] = 0 \,\forall\, \theta \Rightarrow h(T) = 0 \; wp.1$$

Completeness has the following consequence which is very useful when deriving hypothesis tests for exponential family distributions.

Theorem 3 (Basu's Theorem). *If T is a complete sufficient statistic for the family of distributions $\mathcal{P}_\theta$ then any ancillary statistic V is independent of T.*

Proof. By ancillarity, $p_C = P(V \in C)$ does not depend on θ for any C. Let $q_C(t) = P(V \in C|T = t)$. Then $E_\theta[q_C(T)] = p_C$, and so by completeness, $q_C(T) = p_C$ wp. 1. This proves the independence of V and T. □

The notions of sufficiency, minimal sufficiency, and completeness tend to feature prominently in any discussion of the exponential family of distributions.

Definition 5 (s-parameter exponential families). *A class of distributions $\mathcal{P}_\theta = \{P_\theta\}$ is an s-parameter exponential family if the densities of the members of the class have the form*

$$p_\theta(x) = \exp\left[\sum_{i=1}^{n} \theta_i T_i(x) - A(\theta)\right] h(x),$$

where $\theta = (\theta_1, \ldots, \theta_s)' \in E_s$. If the θ_i are linearly independent and the T_i are also linearly independent then the family is full-rank exponential.

[2] Observe that the transformation $h(T)$ cannot contain more "material" than is already contained in T. The key idea is the possibility of using h to extract any ancillary material from T.

It is a useful exercise to check if a commonly-occurring class of distributions is an exponential family.[3]

Here are two useful facts.

Theorem 4. *If $p_\theta(x)$ is the density of a full-rank s-parameter exponential family, then $T(x) = (T_1(x), \ldots, T_s(x))'$ is minimal sufficient.*

Proof. T is obviously sufficient by the factorization criterion. Consider $s+1$ distributions with densities $p(x|\theta^{(j)})$, $\theta^{(j)} = (\theta_1^{(j)}, \ldots, \theta_s^{(j)})'$, $j = 0, 1, \ldots, s$. By Theorem 2 the minimal sufficient statistic is equivalent to

$$\left(\sum_{i=1}^{s}(\theta_i^{(1)} - \theta_i^{(0)})T_i(X), \ldots, \sum_{i=1}^{s}(\theta_i^{(s)} - \theta_i^{(0)})T_i(X)\right)',$$

which is equivalent to $T(X) = (T_1(X), \ldots, T_s(X))'$, given that no nontrivial linear combination of the $\theta_i^{(j)} - \theta_i^{(0)}$ is zero for each $j = 0, 1, \ldots, s$. □

In an exponential family, it turns out that not only is the statistic T minimal sufficient, it is complete.

Theorem 5. *If X is distributed according to a full-rank exponential family with minimal sufficient statistic $T(X) = (T_1(X), \ldots, T_s(X))'$, then T is complete.*

Proof. See Lehmann (1986), Section 4.3, Theorem 1. □

Corollary 5.1. *Let the density of X be given by*

$$p_{\theta,\eta}(x) = C(\theta,\eta)\exp\left[\theta U(x) + \sum_{i=1}^{s}\eta_i T_i(x)\right].$$

For θ equal to some fixed value, consider a statistic V that is ancillary for η. Then V is independent of T for all η.

References

[1] Lehmann, E. L. (1986): *Testing Statistical Hypotheses*, Second Edition. New York: Springer-Verlag.

[2] Lehmann, E. L. and G. Casella (1998): *Theory of Point Estimation*, Second Edition. New York: Springer-Verlag.

[3] Chances are that it will be. Note that the gamma, chi-squared, beta, Bernoulli, binomial, Poisson, and normal classes are all exponential families.