Department of Electrical & Computer Engineering
University of California, Santa Barbara

ECE 240A
Winter 2010
Shynk
H.O. #14

# SUFFICIENT STATISTICS AND UMVU ESTIMATION

- **DEFINITION 1 Statistic:** A *statistic* $T$ is a function of observable random variables (measurements) which is itself an observable random variable, and does not contain any unknown parameter $\theta$. A statistic is a *summary* of the data from which we can make inferences about the joint density of the random variables.

- **DEFINITION 2 Sufficient Statistic:** A statistic $T(\mathbf{z})$ is defined to be a *sufficient statistic* if and only if the conditional distribution of $\mathbf{z}$ given that $T(\mathbf{z}) = t$ does not depend on $\theta$ for any value $t$. Note that $\mathbf{z}$ contains the measurements, i.e., $\mathbf{z} = [z(1), \ldots, z(N)]^T$. A sufficient statistic *condenses* the data in such a way that *no* information about $\theta$ is lost.

- **DEFINITION 3 Jointly Sufficient Statistics:** The statistics $T_1(\mathbf{z}), \ldots, T_r(\mathbf{z})$ are defined to be *jointly sufficient* if and only if the conditional distribution of $\mathbf{z}$ given that $T_1 = t_1, \ldots, T_r = t_r$ does not depend on $\theta$. Note that the components of $\mathbf{z}$ are always jointly sufficient (i.e., the conditional distribution of the sample given the sample does not depend on $\theta$).

- **THEOREM 1 Factorization Theorem:** A statistic $T(\mathbf{z})$ is sufficient if and only if the joint density function of $\mathbf{z}$ factors as

$$f(\mathbf{z}; \theta) = g(T(\mathbf{z}); \theta) h(\mathbf{z})$$

  where the function $h(\cdot)$ is nonnegative and does not involve the parameter $\theta$, and the function $g(\cdot)$ is nonnegative and depends on $\mathbf{z}$ only through the statistic $T(\cdot)$. Note that there are many possible sets of sufficient statistics, and that $h(\cdot)$ may be a constant independent of the data $\mathbf{z}$. This theorem can be used to *find* sufficient statistics, and it is often easier to use than Definition 2. In many cases, the resulting sufficient statistics are minimal and even complete (but not always).

- **LEMMA 1:** If $T$ is a sufficient statistic and $T = H(U)$ where $H(\cdot)$ is some function, then $U$ is also sufficient. Clearly, knowledge of $U$ implies knowledge of $T$. Note that $T$ provides a greater reduction of the data unless $H$ is 1:1, in which case $T$ and $U$ are equivalent.

- **DEFINITION 4 Minimal Sufficient Statistic:** A set of jointly sufficient statistics is defined to be *minimal sufficient* if and only if it is a function of *every* other set of sufficient statistics. Using the notation of Lemma 1, a minimal sufficient statistic can be defined as follows: For *any* sufficient statistic $U$ there exists a function $H$ such that $T = H(U)$. Of all the sufficient statistics, a minimal sufficient statistic provides the greatest possible reduction of the data in terms of the dimensionality. For example, a minimal sufficient statistic may have dimension 1 whereas the original sample may have dimension $N$.

- **DEFINITION 5 Support:** The *support* of a distribution $f(x)$ is the set of all points $x$ for which $f(x) > 0$ (strictly positive). The support of a distribution is important when dividing distributions, such as the likelihood ratio.

- **DEFINITION 6 Family of Distributions:** A *family of distributions* is defined as the set of distributions (possibly uncountable) obtained by varying the parameter $\theta$ over its entire range of values. For example, the set of normal distributions $N(\mu, \sigma^2)$ with $\sigma^2$ known and $\mu$ unknown is a family of distributions.

- **DEFINITION 7 Exponential Family:** A family of densities $f(\mathbf{z};\theta)$ that can be expressed as
$$f(\mathbf{z};\theta) = a(\theta)b(\mathbf{z})\exp\left(\sum_{k=1}^{n} c_k(\theta)d_k(\mathbf{z})\right)$$
for all $\theta$, and for a suitable choice of functions $a(\cdot)$, $b(\cdot)$, $c_k(\cdot)$, and $d_k(\cdot)$ is defined to belong to the *exponential family* or *exponential class*. Examples include the gamma, chi-square, beta, binomial, Poisson, normal, and negative binomial distributions. The uniform distribution is *not* an exponential family. Exponential families have the property that there exists a sufficient statistic of *fixed* size, regardless of the sample size $N$. This property is not always shared by other families of distributions.

- **LEMMA 2:** A necessary and sufficient condition for a statistic $U$ to be sufficient is that for any fixed $\theta$ and $\theta_0$, the ratio $T = f(\mathbf{z};\theta)/f(\mathbf{z};\theta_0)$ is a function of $U(\mathbf{z})$. This result states that $U$ is a sufficient statistic if and only if $T$ is a function of $U$, and it follows from the factorization theorem.

- **THEOREM 2:** Let $F$ be a finite family with densities $f(\mathbf{z};\theta_i)$, $i = 0, \ldots, n$, all having the same support. Then the statistic
$$T(\mathbf{z}) = \left(\frac{f(\mathbf{z};\theta_1)}{f(\mathbf{z};\theta_0)}, \ldots, \frac{f(\mathbf{z};\theta_n)}{f(\mathbf{z};\theta_0)}\right)$$
is *minimal* sufficient. Note that the dimension of $T$ may be less than $n$ since some of the ratios may lead to the same statistic.

- **COROLLARY 1:** Suppose $F$ is a family of distributions with common support and that $F_0$ is a subset of $F$. If $T$ is minimal sufficient for $F_0$ and sufficient for $F$, then it is also minimal sufficient for $F$.

- **DEFINITION 8 UMVU Estimator:** An estimator $T^*(\mathbf{z})$ of $\theta$ is defined to be a *uniformly minimum-variance unbiased estimator* (UMVUE) of $\theta$ if and only if (i) $E(T^*) = \theta$, and (ii) $\text{var}(T^*) \leq \text{var}(T)$ for any other estimator $T(\mathbf{z})$ of $\theta$ which satisfies $E(T) = \theta$.

- **THEOREM 3 Rao-Blackwell:** Let $T(\mathbf{z})$ be a *sufficient* statistic, and let the statistic $T'(\mathbf{z})$ be an unbiased estimator of $\theta$. Define $T^*(\mathbf{z})$ by the *conditional expectation* $T^* = E[T'|T(\mathbf{z})]$. Then, (i) $T^*$ is a sufficient statistic and it is a function of $T(\mathbf{z})$, (ii) $E(T^*) = \theta$, and (iii) $\text{var}(T^*) \leq \text{var}(T')$ for every $\theta$, and $\text{var}(T^*) < \text{var}(T')$ for some $\theta$ unless $T^*$ is equal to $T'$ with probability one. This conditioning may improve the estimate, but it does not guarantee that we will obtain the UMVU estimate – this requires that the statistic $T$ be *complete*.

- **DEFINITION 9 Ancillary:** A statistic $U(\mathbf{z})$ is said to be *ancillary* if its distribution does not depend on $\theta$ and *first-order ancillary* if its expectation $E[U(\mathbf{z})]$ is constant, independent of $\theta$. An ancillary statistic by itself contains no information about $\theta$. Note that a minimal sufficient statistic may contain such ancillary information.

- **DEFINITION 10 Complete Sufficient Statistic:** A sufficient statistic $T$ that satisfies the following property is said to be *complete:*

$$E[g(T)] = 0 \;\text{ for all } \theta \;\text{ implies }\; g(T) = 0.$$

  A sufficient statistic $T$ appears to be most successful in reducing the data if *no* nonconstant function of $T$ is ancillary or even first-order ancillary, i.e., $E[g(T)] = c$ for all $\theta$ implies $g(T) = c$ where $c$ is a constant. The above definition is obtained by subtracting $c$. Another way of stating that a statistic $T$ is complete is the following: $T$ is complete if and only if the *only* unbiased estimator of zero which is a function of $T$ is the statistic that is identically zero with probability one.

- **THEOREM 4 Lehmann-Scheffe:** Let $T(\mathbf{z})$ be a *complete* sufficient statistic for $\theta$, and let $T'(\mathbf{z})$ be any unbiased estimator of $\theta$. Then $T^* = E[T'|T(\mathbf{z})]$ is a UMVU estimator of $\theta$. Note that this is similar to the Rao-Blackwell theorem except that we are now starting with a complete sufficient statistic (instead of only a sufficient statistic). This theorem results in two methods that can be used to find UMVU estimators.