EE7620 Sufficient Statistics Fall 2006

As discussed previously, we are often encountered with the problem of having to estimate something (a random variable M) from the observation of something else (another random variable R). It turns out however that in many cases the data provided by the observation is more than necessary. In other words the observation contains information that is useless as far as the estimation process is concerned. In such situations we often process the observed data by removing the unnecessary information before implementing an estimation rule. Obviously in such processing we do not want to discard any useful information. In the following we will represent this processing by a function. An important question then is what type of functions can be used on the observation so that while the unnecessary data in the observation is removed, no useful information is lost. In the following these notions will be made clear rigorously.

Definition 1: Let f be a given function. $Z = f(R)$ is called a sufficient statistic for decisions about M based on R if M and R are conditionally independent given Z, i.e.,

$$p_{MR|Z}(m, r|z) = p_{M|Z}(m|z)\ p_{R|Z}(r|z) \qquad \text{for all} \quad m, r, z.$$

Note

1. If M and R are independent it does not imply that M and R are conditionally independent given Z. The following example illustrates this fact. Let $\Omega_M = \Omega_R = \{0, 1\}$, and let

   $$p_M(0) = p_R(0) = \frac{1}{2},$$

   Assume that M and R are independent. Let $Z = M \otimes R$, where $\otimes$ denotes modulo 2 addition. Now M and R are independent but they are *not* conditionally independent given Z.

2. If M and R are conditionally independent given Z it does not imply that M and R are independent.

By the chain rule we can write:

$$p_{MR|Z}(m, r|z) = p_{M|Z}(m|z)\ p_{R|MZ}(r|m, z) = p_{R|Z}(r|z)\ p_{M|RZ}(m|r, z)$$

So an equivalent definition of sufficient statistics would be:

Definition 2: $Z = f(R)$ is a sufficient statistic for M based on R if

$$p_{R|MZ}(r|m, z) = p_{R|Z}(r|z) \ \text{ for all } \ m, r, z \ \text{ such that } \ p_{M|Z}(m|z) > 0,$$

or equivalently if

$$p_{M|RZ}(m|r, z) = p_{M|Z}(m|z) \ \text{ for all } \ m, r, z \ \text{ such that } \ p_{R|Z}(r|z) > 0.$$

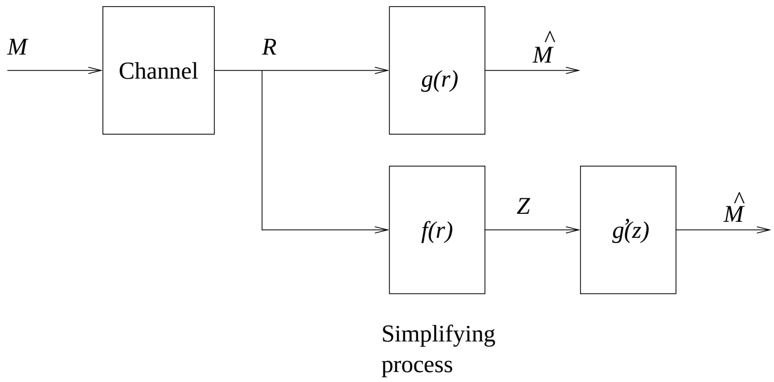

Figure 1:

The following result provides the justification for the definition of sufficient statistic.

Theorem: If Z is a sufficient statistic for decisions about M based on R, then the best decision rule based on Z is as good as the best decision rule based on R.

Proof

To prove this we show the following.

1. The best estimate based on Z and R is as good as the best estimate based on Z.
2. The best estimate based on Z is as good as the best estimate based on Z and R.

Then 1 and 2 prove our claim.

1. This is obvious because we can just ignore R.
2. Proof of 2.

   The best rule based on Z and R is, say, g_1, i.e., Given $(Z, R) = (z, r)$, let

$$g_1(z, r) = m_i \ \text{ iff } \ (z, r) \in I_i$$

where

$$I_i = \{(z, r) : p_{M|ZR}(m_i|z, r) \geq p_{M|ZR}(m_j|z, r) \ \text{ for all } \ j \neq i\}$$

or

$$g_1(z, r) = m_i \ \text{ iff}$$

$$p_{M|ZR}(m_i|z, r) \geq p_{M|ZR}(m_j|z, r) \ \text{ for all } \ j \neq i$$

The best rule based on Z is, say, g_2, i.e., Given $Z = z$, let

$$g_2(z) = m_i \ \text{ iff}$$

$$p_{M|Z}(m_i|z) \geq p_{M|Z}(m_j|z) \ \text{ for all } \ j \neq i$$

We will show that $P_{g_1}(E) = P_{g_2}(E)$ or equivalently $P_{g_1}(C) = P_{g_2}(C)$

Suppose R is discrete. The case when R is continuous is similar.

$$P_{g_1}(C) = \sum_{\Omega_Z \Omega_R} p_{M|ZR}[g_1(z,r)|z,r]\; p_{ZR}(z,r)$$

$$= \sum_{\Omega_Z \Omega_R} [\max_i p_{M|ZR}(m_i|z,r)]\; p_{ZR}(z,r)$$

Now Z is sufficient statistic. Therefore,

$$p_{M|ZR}(m|z,r) = p_{M|Z}(m|z) \;\text{ for all }\; m,r,z \;\text{ such that }\; p_{R|Z}(r|z) > 0.$$

Also if

$$p_{R|Z}(r|z) = 0 \;\text{ then }\; p_{RZ}(r,z) = 0.$$

Thus,

$$\begin{aligned}
P_{g_1}(C) &= \sum_{\Omega_Z,\Omega_R} [\max_i p_{M|Z}(m_i|z)]\; p_{ZR}(z,r) \\
&= \sum_{\Omega_Z,\Omega_R} [\max_i p_{M|Z}(m_i|z)]\; p_{ZR}(z,r) \\
&= \sum_{\Omega_Z} [\max_i \; p_{M|Z}(m_i|z)]\; p_Z(z) \qquad (1) \\
&= P_{g_2}(C).
\end{aligned}$$

This completes the proof.

Examples:

1. Let $Z = f(R)$ where f is a one to one function.

   Claim: Z is a sufficient statistic for M based on R.

   $$p_{R|MZ}(r|m,z) = \begin{cases} 1 & \text{if } \; r = f^{-1}(z) \\ 0 & \text{if } \; r \neq f^{-1}(z) \end{cases}$$

   Also

   $$p_{R|Z}(r|z) = \begin{cases} 1 & \text{if } \; r = f^{-1}(z) \\ 0 & \text{if } \; r \neq f^{-1}(z) \end{cases}$$

   This proves our claim.

2. Consider the following figure where N_1, N_2 and M are independent.

   Let $Z = f(R_1, R_2) = R_1$.

   Claim: Z is a sufficient statistic for M based on (R_1, R_2).

   Proof

We have

$$p_{R_1R_2|MZ}(r_1, r_2|m, z) = p_{R_2|MZ}(r_2|m, z)\ p_{R_1|R_2MZ}\ (r_1|r_2, m, z)$$

Now

$$p_{R_2|MZ}(r_2|m, z) = p_{N_2}(r_2)$$

and

$$p_{R_1|R_2MZ}\ (r_1|r_2, m, z) = \begin{cases} 1 & r_1 = z \\ 0 & r_1 \neq z \end{cases}$$

Therefore,

$$p_{R_1R_2|MZ}(r_1, r_2|m, z) = \begin{cases} p_{N_2}(r_2) & r_1 = z \\ 0 & r_1 \neq z \end{cases}$$

On the other hand,

$$p_{R_1R_2|Z}(r_1, r_2|z) = p_{R_2|R_1Z}(r_2|r_1, z)\ p_{R_1|Z}(r_1|z)$$

Now

$$p_{R_1|Z}(r_1|z) = \begin{cases} 1 & r_1 = z \\ 0 & r_1 \neq z \end{cases}$$

and

$$p_{R_2|R_1Z}(r_2|r_1, z) = p_{N_2}(r_2)$$

Therefore,

$$p_{R_1R_2|Z}(r_1, r_2|z) = \begin{cases} p_{N_2}(r_2) & r_1 = z \\ 0 & r_1 \neq z \end{cases}$$

This proves our claim. In this example we might think of N_2 as irrelevant or unnecessary data that can be ignored without compromising the optimality of the decision rule. Note that in this example our claim will not be true if for instance N_1 and N_2 are not independent random variables.

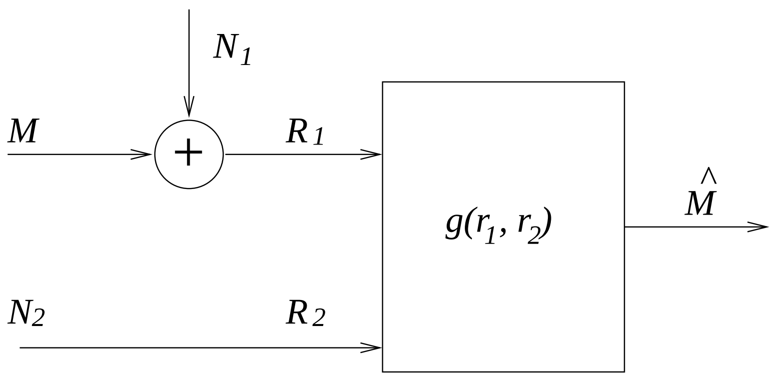

Figure 2: