EE7620 Decision and Estimation Theory

If M is the outcome of some random experiment we would like to know but can not observe, and if R is the outcome of some random experiment we can observe, and if M and R are not independent, then it is reasonable to expect that R might profitably be used to make an estimate of the unobserved value of M. The study of rules for doing just this is the topic of decision and estimation theory. This theory tells us how to find a decision (or estimation) rule which, for any $R = r$ that might occur, indicates a good guess for M. That is, an estimation rule is a function which assigns an estimate $\hat{m}$ to every potential observed value of $R = r$. In particular the theory can provide us with the best such rule.

In order to characterize the best rule we must know the relationship between M and R, i.e., their joint distribution. In addition we must specify how to measure the "goodness" of an estimation rule. Indeed, for different measures of goodness different rules turn out to be the best.

Suppose we have two random variables (r.v.) M and R. M is discrete with alphabet $\Omega_M = \{m_1, m_2, ..., m_q\}$, while R can be discrete or continuous and has alphabet Ω_R. The joint distribution of M and R is given by $p_{MR}(m, r)$. If R is discrete then $p_{MR}(m, r)$ is the probability mass function (pmf), while if R is continuous, $p_{MR}(m, r)$ is mixed. We want to know the value of M. But we can only observe the outcome of the r.v. R and must estimate M from the observation of R. When the r.v. M is discrete, the estimation rule is commonly called a decision rule rather than an estimation rule.

We need a decision rule $g : \Omega_R \longrightarrow \Omega_M$. If the outcome of R is r, then our estimate of M is $g(r)$.

Note that g is a function from Ω_R into Ω_M. Recall the definition of a function. To every element of Ω_R, the function g must assign one and only one element of Ω_M. Why? Because

1. *For every $r \in \Omega_R$ we need a decision for M. We can not have a decision rule that gives up for certain values of R.*
2. *We can not have a situation where for one r we have two decisions.*

Decision Regions

A good way to picture a decision rule is via its *decision regions.* For $i = 1, 2, ..., q$ let $I_i = \{r : g(r) = m_i\}$. I_i is called the i^{th} decision region ($I_i \subset \Omega_R$). If $r \in I_j$, then $g(r) = m_j$ is our estimate of M. The collection of decision regions $\{I_1, I_2, ..., I_q\}$ forms a partition of Ω_R; that is

$$I_i \bigcup I_j = \phi \ \text{ for all } \ i \neq j$$

and

$$\bigcup_{i=1}^{q} I_i = \Omega_R$$

The Minimum Probability of Error Decision Rule

Let us agree that the best decision rule is one that minimizes the probability that it is in error. For a rule $g(r)$ this is defined as

$$\text{Probability of Error} = P(E) = P(M \neq g(R)).$$

We can write

$$P(E) = P(\{(m,r) : m \neq g(r)\}) = \sum_{m_i} \int_{\{r: g(r) \neq m_i\}} p_{MR}(m,r)\ dr = \sum_{m_i} \int_{I_i^c} p_{MR}(m,r)\ dr$$

We might equivalently judge a decision rule by the probability that it is correct given by $P(C) = 1 - P(E)$. We have

$$P(C) = P(M = g(R)) = P(\{(m,r) : m = g(r)\}) = \sum_{m_i} \int_{I_i} p_{MR}(m,r)\ dr.$$

Now when R is a continuous r.v.,

$$P(C) = P(M = g(R)) = \int_{\Omega_R} P(M = g(r)|R = r)\ p_R(r)\ dr$$

The best decision rule is that which maximizes the probability of correct. Now since all the terms in the expression for $P(C)$ are non-negative, we can maximize $P(C)$ by choosing (for each given r) $g(r)$ to be that $m \in \Omega_M$ for which $P(M = m|R = r)$ is maximum.

Example:

Suppose for $R = \alpha$ we have

$$P(M = m_1|R = \alpha) = .1$$

$$P(M = m_2|R = \alpha) = .5$$

and

$$P(M = m_i|R = \alpha) < .5 \ \text{ for } \ i = 3, 4, ..., q$$

So choose $g(\alpha) = m_2$ because

$$P(M = m_2|R = \alpha) > P(M = m_i|R = \alpha) \ \text{ for all } \ i \neq 1$$

Therefore, to maximize $P(C)$, for each $r \in \Omega_R$ let $g(r)$ be that value in Ω_M, say $\hat{m}$, for which $P(M = \hat{m}|R = r)$ is largest. Note that $P(M = \hat{m}|R = r) = p_{M|R}(\hat{m}|r)$.

Therefore, the minimum probability of error decision rule is given by:

Given that $R = r$, let

$$g(r) = m_i \ \text{ if and only if } \ p_{M|R}(m_i|r) \geq p_{M|R}(m_j|r) \quad \text{for all } \ j \neq i$$

If $p_{M|R}(m_k|r) = p_{M|R}(m_l|r) > p_{M|R}(m_j|r)$ for all $j \neq k, l$, then choose $g(r) = m_k$ or $g(r) = m_l$ arbitrarily. The probability of correct (or the probability of error) will be the same for both choices. The above decision rule is called the *Maximum A Posteriori rule (MAP).*

We will now give some other useful expressions for $P(E)$ and $P(C)$.

$$\begin{aligned} P(E) &= P[M \neq g(R)] \\ &= \sum_{m_i} P[g(R) \neq M | M = m_i] \; p_M(m_i) \\ &= \sum_{m_i} P[g(R) \neq m_i | M = m_i] \; p_M(m_i) \end{aligned}$$

Let

$$\begin{aligned} P(E|m_i) &= P(g(R) \neq m_i | M = m_i) \\ &= \int_{r: g(r) \neq m_i} p_{R|M}(r|m_i) \; dr \\ &= \int_{I_i^c} p_{R|M}(r|m_i) \; dr \end{aligned}$$

Then

$$P(E) = \sum_{m_i} P(E|m_i) \; p_M(m_i)$$

Also $P(C) = \sum_{m_i} P(C|m_i) \; p_M(m_i)$ where

$$P(C|m_i) = P[g(R) = m_i | M = m_i] = \int_{I_i} p_{R|M}(r|m_i) \; dr$$

Example:

Let $\Omega_M = \{m_1, m_2\}$ and let $p_M(m_1) = 1/4$. Also let

$$p_{R|M}(r|m_1) = \begin{cases} e^{-r} & r \geq 0 \\ 0 & r < 0. \end{cases}$$

and

$$p_{R|M}(r|m_2) = \begin{cases} 2e^{-2r} & r \geq 0 \\ 0 & r < 0. \end{cases}$$

Find the best decision rule and $P(E)$.

Solution:

We need $p_{M|R}(m_i|r)$. We can obtain this from the Bays' rule:

$$p_{M|R}(m_i|r) = \frac{p_{R|M}(r|m_i) \; p_M(m_i)}{p_R(r)}$$

Best decision rule:

$$g(r) = m_i \;\; \text{if and only if}$$

$$\frac{p_{R|M}(r|m_i) \; p_M(m_i)}{p_R(r)} \geq \frac{p_{R|M}(r|m_j) \; p_M(m_j)}{p_R(r)} \;\; \text{for all} \;\; j \neq i$$

Therefore,

$$g(r) = m_1 \;\; \text{if and only if} \;\; p_{R|M}(r|m_1) \; p_M(m_1) \geq p_{R|M}(r|m_2) \; p_M(m_2)$$

or

$$\frac{1}{4}e^{-r} \geq \frac{3}{4}2e^{-2r}$$

Thus

$$g(r) = m_1 \quad \text{if and only if} \quad r \geq \ln 6$$

$$I_1 = \{r : r \geq \ln 6\}, \quad I_2 = I_1^c$$

$$g(r) = \begin{cases} m_1 & r \geq \ln 6 \\ m_2 & r < \ln 6 \end{cases}$$

$$P(E) = \sum_{m_i} P(E|m_i)p_M(m_i)$$

$$P(E|m_1) = \int_{r:r\in I_1^c} p_{R|M}(r|m_1)\ dr = \int_0^{\ln 6} e^{-r}\ dr = 1 - e^{-\ln 6} = \frac{5}{6}$$

and

$$P(E|m_2) = \int_{r:r\in I_2^c} p_{R|M}(r|m_2)\ dr = \int_{\ln 6}^{\infty} 2e^{-2r}\ dr = e^{-2\ln 6} = \frac{1}{36}$$

$$P(E) = \frac{5}{6} \times \frac{1}{4} + \frac{1}{36} \times \frac{3}{4} = \frac{11}{48}$$

Thus $P(C) = \frac{37}{48}$.

Figure 1:

Maximum Likelihood Decision Rule

This is a special case of the MAP rule. The MAP rule is given by

Given observation $R = r$, $g(r) = m_i$ if and only if

$$p_{R|M}(r|m_i)\ p_M(m_i) \geq p_{R|M}(r|m_j)\ p_M(m_j) \ \text{ for all } \ j \neq i$$

Now suppose $p_M(m_i) = \frac{1}{q}$ for $i = 1, 2, ..., q$, i.e., all the values of M are equally likely. Then the MAP rule simplifies to

Given $R = r$, Let $g(r) = m_i$ if and only if

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

This is called the Maximum Likelihood rule (ML). For $R = r$, $p_{R|M}(r|m)$ is the likelihood of m.