Journal of Economic Theory and Econometrics, Vol. 25, No. 4, Dec. 2014, 71–81

Optimal Aggregation of Multiple Signals: Optimality of Linear Aggregation Rule and Possibility of Using the Maximum or the Minimum of Signals

Sam-Ho Lee* Joonbae Lee†

Abstract An optimal aggregation rule of multiple signals is studied. A sufficient condition for the optimality of a linear aggregation rule (the weighted sum of signals) and the common classes of signal distributions satisfying the condition are provided. Other aggregation rules such as rules using the maximum or the minimum are also considered and their usefulness is illustrated through examples.

Keywords Signal Aggregation, Linear Aggregation, Rule using the Maximum or the Minimum

JEL Classification D89

*Department of Economics, Korea University, Email: samho@korea.ac.kr
†Department of Economics, University of Pennsylvania, Email: joonbae@sas.upenn.edu

Received August 5, 2014, Revised November 12, 2014, Accepted December 9, 2014

1. INTRODUCTION

When evaluating an individual's performance, we may have multiple signals available. For instance, in university admission, universities can use SAT scores, high school GPA, extra-curricular activities, recommendation letters, or essays. Universities attempt to infer an applicant's underlying character from these measures. When firms are hiring or promoting employees, they also have multiple signals at hand such as sales performance, peer review, and test scores.

When multiple signals are available, we tend to use the weighted sum of all signals in evaluation. This practice is partially justified in that each measure represents different aspects of the underlying quality and we wish to obtain a complete picture of the individual. But some signals may represent the same dimension of quality, and the weighted sum is not the only way of aggregating multiple signals. For example, SAT scores and high school GPA may represent the same underlying academic capacity, and we can use the maximum or the minimum of two signals instead of the weighted sum to select who is admitted.

This paper investigates the optimal way of aggregating multiple signals that represent the same dimension of an underlying quality. By investigating this optimal rule, it will be determined whether the weighted sum rule can be justified and, if so, in what environment. Some existing aggregation rules using the maximum or the minimum of the signals will be also considered. This, although appearing somewhat unusual, is a rule that is quite often used. For the admission of international students to U.S. academic institutions, a TOEFL score is needed, and applicants can choose which score to have sent to the institution to which they apply. This practice leads to the consideration of the maximum score only.[1] Korean universities have several separate admissions processes, and among them one mainly focuses on high school GPA while another focuses on the CSAT (College Scholastic Ability Test, equivalent to the SAT). Under this system, a student will be admitted if he or she passes the relevant threshold in either of the two components; in other words, universities use the maximum score from the two components. In some tests, one fails if one does not meet the minimum standard in any subject, regardless of the total score. In that case, the minimum scores of many subjects are used for selection. We will also examine whether these seemingly unusual rules can be justified.

The optimal aggregation rule is, of course, dependent on the distributional characteristics of the signal. The aggregation rule implicitly compares two types

[1] GRE score reporting, on the other hand, requires that all the scores received under the same test name be reported.

of individuals if we have two signals available: one with two middle signals and one with a high signal and a low signal. While the (weighted) sum rule (or linear rule) treats these two types similarly, the rule using the minimum favors the former and the rule using the maximum favors the latter. This comparison should depend on the distributional character of the signals. This paper shows that linear aggregation, such as weighted sum, is the optimal rule for a very general class of signal distributions, which justifies the wide use of it. We also illustrate, however, that there exists a common distribution with which a rule using the maximum or the minimum outperforms the linear rule. Therefore, we can accommodate the existence of such rules.

Information aggregation is an important issue in statistical inference. As sample observation carries information about population (or more specifically population parameters), it is an essential question how to aggregate sample observations to estimate population parameters. Our question is a subset of the statistical inference problem and, as is shown later, the optimality of linear aggregate of the signals in the general class of distribution, though derived through a different method, is a corollary of a well-known concept of sufficient statistic.[2] Our framework, however, also allows us to investigate the possibility of better performance of the rule using the maximum or the minimum of the signals.

The concept of the sufficient statistic is also used to evaluate the informational value of the signal in a principal-agent setting (Holmstrom 1979).[3] The optimality of the linear aggregate of the signals in the same setting is discussed in Banker and Datar (1989). Banker and Datar consider the optimality of linear aggregate in a principal-agent setting where two observable signals of an agent's action are available. In a usual principal-agent setting, compensation is determined as a function of these two observable signals for motivating agents. Banker and Datar decompose this compensation function into two steps: an aggregation of two signals and compensation dependent upon this aggregate. They provide a sufficient condition on the joint density function of signals for the optimality of a linear aggregate in the first step.[4] This paper is also partly concerned about the optimality of a linear aggregate but in the statistical inference context.

[2]See Blackwell and Girshick (1954). We thank an anoymous referee for indicating this and introducing the relevant reference.

[3]See also Kim (1995) and Jewitt (1997) for further discussion on information issues in principal-agent set up.

[4]Their sufficient condition for the optimality of a linear aggregate is also related to the concept of sufficient statistic though it was not explicitly discussed there. The proposed functional form of the joint density function of the singals guarantees that the linear aggregate of the signal is a sufficient statistic of the agent's chosen action.

This paper is organized as follows: Section 2 introduces the model setting. In Section 3 we show how the optimal rule is obtained and that the optimal rule is a linear rule for a general class of signal distribution. Better performance of the aggregation rules using the maximum or the minimum is also illustrated in that section. The conclusion follows.

2. MODEL

2.1. POPULATION AND SELECTION

There is a unit mass of agents with two types, q mass of high types (H) and $1-q$ mass of low types (L). An evaluator wants to select a portion $p(<1)$ of the population and tries to maximize the mass of type H among those selected.

2.2. INFORMATION

The evaluator does not observe agents' types, but instead observes signals of types. There are two signals available, s_1 and s_2, which are distributed conditional on types. We assume the two signals are iid with common support $[\underline{s},\overline{s}]$ with a continuously differentiable conditional density function $f(\cdot|\cdot)$. Furthermore, the signals are informative of types satisfying the (strict) monotone likelihood ratio property, or MLRP,

$$\frac{f(s|H)}{f(s|L)} \leq (<) \frac{f(s'|H)}{f(s'|L)} \quad \text{if } s < s'.$$

As is clear in the later analysis, strict MLRP is necessary to avoid the random selection rule. Without any specific comment, hereafter, we will assume that strict MLRP holds.

2.3. SELECTION PROBLEM

Let g be a selection rule. It is a function which assigns a probability of being selected when a pair of signals (s_1,s_2) is given; $g:[\underline{s},\overline{s}]^2 \to [0,1]$. An agent with signal (s_1,s_2) is selected with probability $g(s_1,s_2)$. A selection rule is feasible if it satisfies the selection capacity p:

$$q\iint_{\underline{s}}^{\overline{s}} g(s_1,s_2) f(s_1|H) f(s_2|H)\, ds_1 ds_2 + (1-q)\iint_{\underline{s}}^{\overline{s}} g(s_1,s_2) f(s_1|L) f(s_2|L)\, ds_1 ds_2 = p. \tag{1}$$

The selection problem is to choose a feasible selection rule g that maximizes the mass of type H among the selected,

$$\max_g \iint_{\underline{s}}^{\overline{s}} g(s_1, s_2) f(s_1|H) f(s_2|H) \, ds_1 ds_2 \tag{2}$$

subject to (1).

3. ANALYSIS

3.1. OPTIMAL SELECTION RULE

The Lagrangian of the problem (2) can be written as[5]

$$\mathscr{L} = \iint_{\underline{s}}^{\overline{s}} \left\{ \begin{array}{c} qg(s_1, s_2) f(s_1|H) f(s_2|H) \\ -\lambda \left[qg(s_1, s_2) f(s_1|H) f(s_2|H) + (1-q) g(s_1, s_2) f(s_1|L) f(s_2|L) \right] \end{array} \right\} ds_1 ds_2 + \lambda p.$$

This is a linear function of $g(s_1, s_2)$ and therefore $g(s_1, s_2)$ takes the value 1 or 0 depending on its coefficient being positive or negative. Thus, the optimal selection rule g^* is

$$g^*(s_1, s_2) = \left\{ \begin{array}{l} 0 \text{ if } L(s_1, s_2) < k \\ 1 \text{ if } L(s_1, s_2) \geq k \end{array} \right., \tag{3}$$

where $L(s_1, s_2) = \frac{f(s_1|H) f(s_2|H)}{f(s_1|L) f(s_2|L)}$, which is a likelihood ratio of two types given the observed pair of signals (s_1, s_2), and $k = \frac{\lambda}{1-\lambda} \frac{1-q}{q}$.

The optimal selection rule is to select agents with signals who are more likely to be type H. Therefore, we rank the pairs of signals by their likelihood ratios and select the agents with pairs from the highest rank until the selection capacity p is fulfilled.

Proposition 1. *The optimal selection rule is given by (3) where k (or λ) is determined to satisfy (1).*

Example 1 (Discrete case)**.** *There is a unit mass of agents with q. Two signals, s_1 and s_2, are conditionally iid and can take three discrete values $\{h, m, l\}$ with $h > m > l$. Conditional probability is given by the following table.*

	H	L
h	$\frac{1}{3}$	$\frac{1}{6}$
m	$\frac{1}{3}$	$\frac{1}{3}$
l	$\frac{1}{3}$	$\frac{1}{2}$

[5] We multiply the objective function by q, as it does not change the solution.

Note that signals are informative in MLRP sense as $\frac{\Pr(h|H)}{\Pr(h|L)} > \frac{\Pr(m|H)}{\Pr(m|L)} > \frac{\Pr(l|H)}{\Pr(l|L)}$. *It is better to select agents with higher signals. The only uncertain comparison is that between* (h,l) *and* (m,m). *If we compare the likelihood ratios of these two pairs,*

$$L(h,l) = L(l,h) = \frac{4}{3}\frac{q}{1-q} > L(m,m) = \frac{q}{1-q}.$$

Then the selection occurs in the following sequence starting with (h,h), *then* $(h,m) = (m,h)$, *then* $(h,l) = (l,h)$, *then* (m,m), *then* $(m,l) = (l,m)$, *and then* (l,l). *The selection stops at the pair when the selection capacity* p *is filled. Since the likelihood ratio is the same for the same signal pair with significant mass, there can be random selection for some capacity parameters* p.

3.2. FROM SELECTION RULE TO AGGREGATION RULE

Until now, only the optimal selection rule has been discussed. But in the introduction, the focus was on how to aggregate two signals with the linear rule as an example. The optimal selection rule is easily translated into an aggregation rule.

The optimal aggregation rule is defined by a likelihood function $L(s_1,s_2)$ at value k. Because of the MLRP of signals, $L(s_1,s_2)$ is an increasing function. Therefore, it is optimal to use $L(s_1,s_2)$ at critical level k as an aggregation rule and to select agents with a pair of signals obtaining a higher score than k. Graphically, $L(s_1,s_2) = k$ is an iso-likelihood curve and the optimal aggregation rule is dependent upon the shape of this curve. Some common aggregation rules are illustrated in Figure 1. If the optimal aggregation rule is to be a (weighted) sum of signals, the iso-likelihood curve should be linear, as in (a). If the minimum or the maximum of signals is to be the optimal aggregation rule, the iso-likelihood curve should be as in (b) or (c).

3.3. THE OPTIMALITY OF A LINEAR AGGREGATION RULE

If the linear rule is to be the optimal aggregation rule, the iso-likelihood curve should be linear regardless of its value. As the shape of an iso-likelihood ratio curve is usually dependent on the level k, it seems difficult to justify using a (weighted) sum as a typical information aggregation rule when there are multiple signals. However, it turns out that the iso-likelihood curve is linear in a very general class of signal distributions. By abusing the notation, let $L(s) = \frac{f(s|H)}{f(s|L)}$ as the likelihood ratio of a single signal. As signals are independent, then

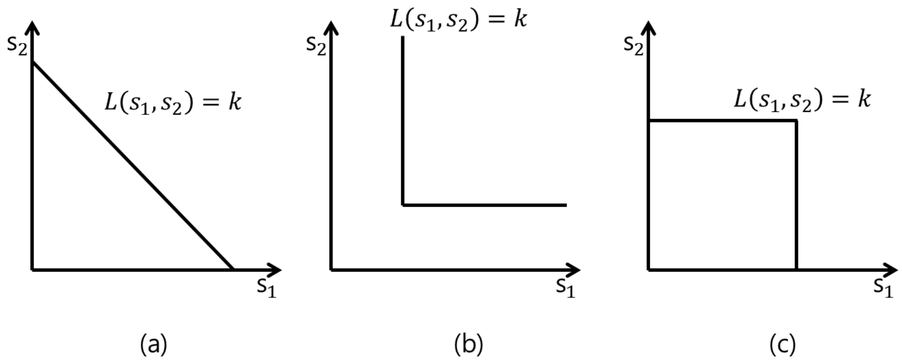

Figure 1: Shape of iso-likelihood curve and optimal aggregation rule

$L(s_1, s_2) = L(s_1)L(s_2)$. The following proposition shows a condition of $L(s)$ for a linear iso-likelihood curve, or the optimality of the linear aggregation rule.

Proposition 2. *Iso-likelihood curves are linear at all levels, and thus a linear aggregation rule is optimal if* $L''(s)L(s) = (L')^2$, *or* $\frac{L'}{L} = \frac{L''}{L'}$.

Proof. Iso-likelihood curves are given by

$$L(s_1, s_2) = L(s_1)L(s_2) = k.$$

Its slope is given as

$$\frac{ds_2}{ds_1} = -\frac{L_1}{L_2} = -\frac{L'(s_1)L(s_2)}{L(s_1)L'(s_2)}.$$

If the slope is constant as s_1 changes, iso-likelihood curves are linear.

$$\frac{d}{ds_1}\frac{ds_2}{ds_1} = -\frac{L''(s_1)L(s_1) - L'(s_1)^2}{L(s_1)^2}\frac{L(s_2)}{L'(s_2)} + \frac{L'(s_1)^2}{L(s_1)^2}\frac{L'(s_2)^2 - L(s_2)L''(s_2)}{L'(s_2)^2}\frac{L(s_2)}{L'(s_2)}$$

Thus, the slope is constant if $L''(s)L(s) = (L')^2$. □

We provide the condition on the likelihood ratio for the optimality of the linear aggregation rule. We need increasing $L(s)$ for the MLRP and $\frac{L'}{L} = \frac{L''}{L'}$ for the optimality of the linear aggregation rule. As is shown in the following remark and examples, there is a large class of common distributions that satisfy the condition.

Remark 1. *1. If a likelihood ratio is log-linear, $L(s) = \kappa \exp(as+b)$, then a linear aggregation rule is optimal. Among the distributions satisfying the MLRP,*[6] *those whose density exhibits such a property include exponential and Laplace, and Normal and gamma with a restriction on the parameters.*
2. If a linear aggregate of signals is a sufficient statistic for types, then it should also be a criterion of the selection problem.[7] *In a statistical inference literature, it is also well known that a linear aggregate is a sufficient statistic in the above mentioned classes of distributions, and thus this result is just a restatement of a sufficiency argument through a different approach.*

Example 2. *i) Exponential distribution has a density $f(s) = \lambda e^{-\lambda s}$. Since the mean of this distribution is $\frac{1}{\lambda}$, we assume that type H has a higher mean, $\lambda_H < \lambda_L$. Then $L(s) = \frac{\lambda_H}{\lambda_L} e^{-(\lambda_H - \lambda_L)s}$ is monotone increasing, and log-linear.*
ii) A similar argument also applies to Laplace distribution, though MLRP holds in a weak sense.[8]
iii) Gamma distribution has a density $f(s) = \frac{1}{\Gamma(m)} x^{m-1} \theta^m e^{-x\theta}$ with $m, \theta > 0$. With m identical for two types and $\theta_H < \theta_L$, $L(s) = (\frac{\theta_H}{\theta_L})^m e^{-x(\theta_H - \theta_L)}$ is increasing in s and log-linear.

Example 3. *When signals are normally distributed and conditional densities $f(s|H)$ and $f(s|L)$ vary only in mean $\mu_H > \mu_L$, but not in variance, the likelihood ratio is monotone increasing and log-linear,*

$$L(s) = \frac{f(s|H)}{f(s|L)} = \exp\left\{ \frac{\mu_H - \mu_L}{\sigma^2} s - \frac{\mu_H^2 - \mu_L^2}{2\sigma^2} \right\}.$$

If two signals have different variances σ_1^2 and σ_2^2, the optimal aggregation rule is the weighted sum and the weight for each signal is the inverse of its variance.

$$L(s_1, s_2) = L(s_1) L(s_2) = \exp\left\{ (\mu_H - \mu_L) \left(\frac{s_1}{\sigma_1^2} + \frac{s_2}{\sigma_2^2} \right) - \left(\mu_H^2 - \mu_L^2\right) \left(\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2} \right) \right\},$$

and iso-likelihood curves are represented by $\frac{s_1}{\sigma_1^2} + \frac{s_2}{\sigma_2^2} = k$.[9]

[6]See Bagnoli and Bergstrom (2004) for a class of distributions satifying MLRP.

[7]A statistic $T(s_1, s_2)$ is sufficient for type if the joint distribution of (s_1, s_2) conditional on $T(s_1, s_2)$ is not dependent on type. Alternatively, it is sufficient if the joint density can be written as $g(T(s_1, s_2), t) h(s_1, s_2)$ where $t = H$ or L by factorization theorem.

[8]For some threshold values k, the iso-likelihood curve can be thick because $L(s)$ is constant for some ranges. Although there can be multiple optimal aggregation rules in this case, the linear rule is still one of them.

[9]It can be also shown that $\frac{s_1}{\sigma_1^2} + \frac{s_2}{\sigma_2^2}$ is a sufficient statistic for μ, and thus that should be a selection criterion.

Since the optimality of linear aggregate in the above mentioned classes of distributions can also be obtained through the sufficient statistic argument, it is readily extended to multiple signals and multiple types. For example, suppose there are I types, $\{t_1, \cdots, t_I\}$ with the mass of each, q_i, and J signals, $\{s_1, \cdots, s_J\}$, which are independently distributed as $s_j \sim N\left(t_i, \sigma_j^2\right)$.[10] Then the weighted sum $\sum_{i=1}^{I} \frac{s_i}{\sigma_i^2}$ is the optimal aggregation rule. The sum $\sum_{i=1}^{I} \frac{s_i}{\sigma_i^2}$ is a sufficient statistic for mean t by factorization theorem, or alternatively likelihood ratio between any two types $L(s, (t_i, t_{i'})) = \frac{q_i}{q_{i'}} \exp\left\{(\mu_i - \mu_{i'}) \sum_{i=1}^{I} \frac{s_i}{\sigma_i^2} - \left(\mu_i^2 - \mu_{i'}^2\right) \sum_{i=1}^{I} \frac{1}{2\sigma_i^2}\right\}$ is ordered by $\sum_{i=1}^{I} \frac{s_i}{\sigma_i^2}$.

3.4. OTHER COMMON RULES OF AGGREGATION

Until now, our analysis has focused on the optimality of linear aggregation rule. However, there are other used rules. As introduced earlier, at one extreme, one can set minimum requirements for all signals, and at the other extreme, one can use only the maximum signal. This section illustrates examples where those rules work better than the linear aggregation rule.

Iso-likelihood curves should be convex (or concave) if a rule using the minimum (or the maximum) of signals is to work better than a linear rule. As seen in Figure 1, a rule using the minimum (or the maximum) represents an extremely concave (or convex) iso-likelihood curve. As a corollary to Proposition 2, we can provide a condition for convex (or concave) iso-likelihood curves.

Corollary 2. *With two signals, iso-likelihood curves are convex (concave) to the origin when likelihoods of both signals exhibit $L''(s)L(s) < (>) L'^2$ for both signals.*

We can find a common distribution that satisfies the above condition, and it turns out that a rule using the maximum (or the minimum) can perform better than a linear rule with that distribution, as illustrated in the following examples.

Example 4. *Assume that the signals are iid Gumbel conditional on types $\{H, L\}$. The Gumbel distribution has distribution and density functions,*

$$F(x|\mu) = e^{-e^{-\frac{x-\mu}{\beta}}} \textit{ and } f(x|\mu) = \frac{1}{\beta} e^{-\frac{x-\mu}{\beta}} e^{-e^{-\frac{x-\mu}{\beta}}}$$

[10] When there are multiple types, the objective of an evaluator is not straightforward as in the model. If there are three types, H, M, and L, then the evaluator may want to maximize the mass of type H, or to minimize that of type L among the selected. In the following argument, however, we do not face the conflict of different objectives as the selection to maximize the mass of type H also minimizes that of type L.

for $\beta > 0$. The mean of the Gumbel distribution is linearly increasing in μ, and the two types are represented by μ_H and μ_L with $\mu_H > \mu_L$. As

$$L(x) = \frac{f(x|H)}{f(x|L)} = e^{\frac{\mu_H - \mu_L}{\beta}} \exp\left(-e^{-\frac{(x-\mu_H)}{\beta}} + e^{-\frac{(x-\mu_L)}{\beta}}\right),$$

$L'(x) > 0$ with $\mu_H > \mu_L$. Furthermore, we can show

$$\frac{L''}{L'} - \frac{L'}{L} = \frac{d}{dx}\left[\ln L'(x) - \ln L(x)\right] = -\frac{1}{\beta} < 0.$$

The iso-likelihood curves are convex to the origin and they become extremely convex as $\beta \to 0$. Therefore, a rule using the minimum performs better than the linear rule for a sufficiently small β.

β	*linear*	*min*
1	1.0000	1.0359
2	1.0000	1.0188
3	1.0000	1.0084
4	1.0000	1.0060
5	1.0000	1.0015

The above table shows the results from an experiment with two independent signals that are Gumbel distributed with $\mu_H = 1$ and $\mu_L = 0$, for $q = 0.5$ and $p = 0.5$. The mass of the selected type H is shown, while the numbers are normalized so that the linear rule has value of 1.

Example 5. *One can think of a mirror example, a negative transformation of the Gumbel distribution,*

$$F(x|\mu) = 1 - e^{-e^{\frac{x+\mu}{\beta}}}, \quad f(x|\mu) = \frac{1}{\beta} e^{\frac{x+\mu}{\beta}} e^{-e^{\frac{x+\mu}{\beta}}}$$

for $\beta > 0$. Suppose type H is represented by a lower μ. Then the MLRP is satisfied. Also, as

$$\frac{L''}{L'} - \frac{L'}{L} = \frac{1}{\beta} > 0,$$

the iso-likelihood curves are concave and become extremely concave as $\beta \to 0$. Therefore, the rule using the maximum can perform better than the linear rule.

β	*linear*	*max*
1	1.0000	1.0345
2	1.0000	1.0156
3	1.0000	1.0152
4	1.0000	1.0115
5	1.0000	1.0108

The above table shows the results from the same experiment in Example 4 except $\mu_H = 0$ *and* $\mu_L = 1$.

4. CONCLUSION

We considered the optimal aggregation rule of multiple signals in a very stylized setting.[11] A sufficient condition on the distribution of signals for the optimal linear aggregation rule is provided. Several common distributions including normal satisfy this condition, though optimality of linear aggregation rule in the same class can be obtained through sufficient statistic argument. This justifies the prevalent use of a linear aggregation rule. We also exemplified the better performance of rules using the maximum or the minimum relative to a linear rule, which are used in some contexts.

REFERENCES

Bagnoli, Mark and Ted Bergstrom. "Log-Concave Probability and Its Applications." Economic Theory 26, 2005, pp. 445-469

Banker, Rajiv and Srikant Datar. "Sensitivity, Precision, and Linear Aggregation of Signals for Performance Evaluation." Journal of Accounting Research, 27(1), 1989, pp. 21-39.

Blackwell, David and M. A. Girshick. Theory of Games and Statistical Decisions. New York: John Wiley & Sons, 1954.

Jewitt, Ian. "Information in Principal Agent Problems." University of Bristol Discussion Paper 96-414, 1996.

Holmstrom, Bengt. "Moral Hazard and Observability." Bell Journal of Economics, 1979, pp. 74-90.

Kim, S. K. "Efficiency of an Information System in an Agency Model." Econometrica, 63(1), 1995, pp. 89-102.

[11] As an anonymous referee indicated, this paper proposes too stylized a model to make a fully developed theoretical paper. This paper is more of an illustrative piece to show the optimality of a seemingly strange aggregation rule; the rule using the maximum or the minimum signal only. The fully developed model is left as a future task.