The Weight of Competence under a Realistic Loss Function

Stephan Hartmann* and Jan Sprenger†

October 16, 2009

Abstract

In many scientific, economic and policy-related problems, pieces of information from different sources have to be aggregated. Typically, the sources are not equally competent. This raises the question of how the relative weights and competences should be related to arrive at an optimal final verdict. Our paper addresses this question under a more realistic perspective of measuring the practical loss implied by an inaccurate verdict.

1 Introduction

When information from different sources is aggregated, be it predictions of scientific models, measurements of different instruments, or opinions of members of a group, it is rarely the case that all sources are equally reliable. Typically, the degree of competence or accuracy varies: some models are known to be more reliable than others, some instruments measure more accurately, some group members possess superior expertise, due to their qualification, knowledge or experience (Lehrer and Wagner 1981).

If we want to obtain an optimal final verdict, we are well advised to take these differences into account. For example, when averaging the predictions of statistical models, the performance of these models with respect to the data is used to determine different relative weights of the models in future predictions (Hoeting et al. 1999).

We transfer this approach to the problem of rational information pooling. A pooling procedure is conceived of as *rational* if it transforms individual pieces of information, together with information about the expertise or accuracy of the sources, in a justifiable way into a final group judgment. More precisely, we investigate the question of how to transform individual competence into relative weight when forming a rational judgment. This question can be applied equally to all problems of opinion pooling where individual contributions are

*To contact the author, write to: Tilburg Center for Logic and Philosophy of Science, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands, e-mail: s.hartmann@uvt.nl, webpage: www.stephanhartmann.org.

†To contact the author, write to: Tilburg Center for Logic and Philosophy of Science, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands, e-mail: j.sprenger@uvt.nl, webpage: www.laeuferpaar.de.

valuable, yet to different degrees, due to the different levels of competence. These problems are pervasive in science, the economy, and in policy-making.

A classical predecessor of our paper is Shapley and Grofman (1984). We replace their approach by a model where the losses suffered by a wrong or imprecise decision are modeled more realistically than in standard statistical theory. In particular, we set up some adequacy conditions on a realistic loss function and propose a peculiar class of functions (section 2). Then, we set up a mathematical model where we map degrees of expertise onto optimal relative weights (section 3). Section 4 summarizes our results and concludes.

2 A Realistic Loss Function

We model the problem of making a sensible final judgment as an *estimation problem*: there is a unknown numerical quantity μ which we would like to estimate, and the individual judgments X_i, $i \leq n$, are modeled as independent random variables that scatter around the true value μ with variance σ_i^2. This approach is inspired by the idea that the information sources resemble measurement instruments with some degree of precision.

The central task consists in finding an estimate $\hat{\mu}(X_1, \ldots, X_n)$ that makes optimal use of the available information. But how shall we evaluate the quality of such an estimator? A standard measure in similar statistical problems is the expected quadratic loss $E[(\hat{\mu} - \mu)^2]$. Then our problem would be the standard problem of finding the ordinary least square estimate, and we could build on an elaborate mathematical theory. But the quadratic loss has severe drawbacks: First, the losses are unbounded whereas in real decisions, there is in general a finite set of options and a worst outcome. Second, large deviations are penalized to a much higher degree than small deviations, due to the convexity of the quadratic function. For example, it is in many situations not clear why a 9% deviation should be nine times as bad as a 3% deviation. Third, for practical purposes it usually does not matter whether one is grossly or very grossly mistaken. This observation has been confirmed experimentally: Kahneman and Tversky (1992, 2000) showed that decision-makers are decreasing sensitivity to large deviations from the true value. But quadratic loss fails to account for this intuition.

We propose the following adequacy conditions on a loss function L:

Smoothness and Boundedness The loss function $L : \mathbb{R}^{\geq 0} \rightarrow [0, 1]$ is an element of $\mathcal{C}^{\infty}$.

Monotony L is monotonously increasing: $L'(x) \geq 0 \; \forall x \geq 0$.

Asymptotic Behavior The loss rate approaches zero for the limiting points: $\lim_{x \to 0} L'(x) = 0$ and $\lim_{x \to \infty} L'(x) = 0$.

These conditions are easily motivated. As argued above, when there is a "worst case", it is reasonable to assume a bounded loss function, and we normalize the range of L to $[0, 1]$. Monotony is self-evident: the more severe the error, the higher the loss. Together this implies the asymptotic behavior of L (concave, decreasing increments). On the other hand, it is plausible that a prediction that is "almost right" is in practice just as good as a fully precise assumption. This justifies the condition $\lim_{x \to 0} L'(x) = 0$, and the behavior of the quadratic

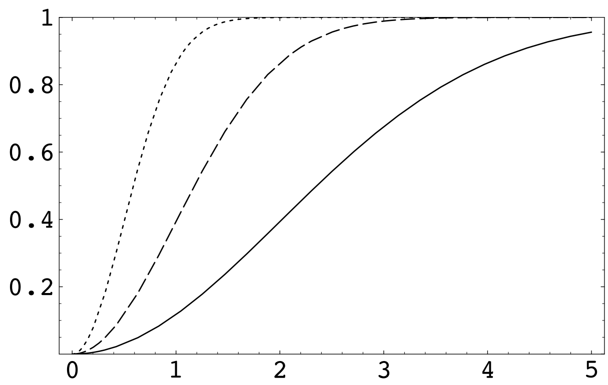

Figure 1: The loss function $L_\alpha(x)$ for $\alpha = .5$ (dotted line), $\alpha = 1$ (dashed line) and $\alpha = 2$ (solid line).

loss function is mimicked for small losses. Finally, by Rolle's theorem, all this implies the existence of an inflection point.

Skewness There is an x_0 such that $\begin{cases} L''(x) > 0 & \text{for all } x < x_0, \\ L''(x) < 0 & \text{for all } x > x_0. \end{cases}$

There are many loss functions which fit that our four adequacy conditions, but we believe that a particularly elegant family of functions is given by

$$L_\alpha(x) = 1 - e^{-\frac{1}{2\alpha^2}x^2}. \tag{1}$$

A further advantage of this family of functions is that it also plays a crucial role in statistical theory. Here α represents the point where the loss rate becomes sublinear. See also figure 1. We contend that these functions are suitable for purposes of decision-making by combining different measurements, predictions, or opinions. Using them instead of the conventional quadratic loss function is an innovation compared to previous approaches of opinion-pooling, and the scale parameter α allows a flexible adaptation of the loss function L_α to the specifics of a particular problem.

3 Expertise and Relative Weight

As mentioned above, the individual judgments of the (not necessarily human) agents are modeled as estimates X_i that scatter around the true value μ. Now, we impose the additional constraint that they scatter symetrically. In particular, the individual estimates are *unbiased*: the agents have no systematic bias towards either a lower or higher value of μ.

At this point, we would like to stress that our paper is intended as a contribution to social epistemology, not to social choice theory. And so considerations of strategic voting, dishonesty or manipulation (e.g. distortion of judgments) have no place: all agents, even if they are human, submit their judgments in the best intention to capture the truth about μ. There is no systematic bias around; error occurs by chance, because one cannot be right all the time.

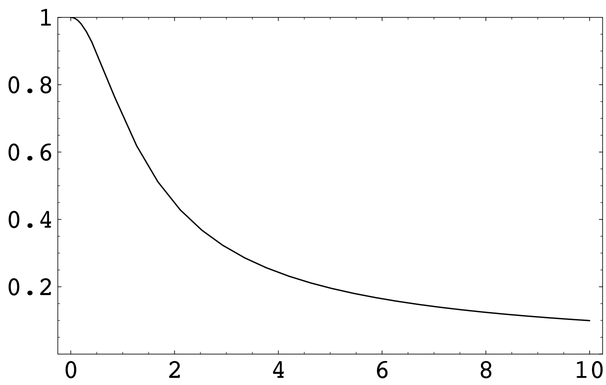

Figure 2: The competence s_i as a function of σ_i/α.

For reasons of convenience (and because we don't see superior modeling alternatives), we assume that the X_i are normally distributed: $X_i \sim N(\mu, \sigma_i)$. Furthermore, we write the group judgment as a linear combination of the individual judgments:

$$\hat{\mu} = \sum_{i=1}^{n} c_i \, X_i, \tag{2}$$

where n denotes group size and the c_i denote individual weights. Now, we ask which values of the c_i minimize the expected loss $E[L_\alpha(\hat{\mu} - \mu)]$ for a given expertise σ_i?

Before we can actually tackle this question, we have to say a word on the σ_i. Obviously, the higher the σ_i, the *lower* an agent's competence. Therefore, we propose to measure individual competence by

$$s_i := E\left[S_\alpha(X_i - \mu)\right], \tag{3}$$

with the *success function* $S_\alpha(x)$ defined by

$$S_\alpha(x) := 1 - L_\alpha(x) = e^{-\frac{1}{2\alpha^2}x^2}. \tag{4}$$

This leads to an inverse relationship between competence and variance. We easily establish the following relationship between both quantities (for a proof, see Appendix A):

$$s_i = \frac{\alpha}{\sqrt{\alpha^2 + \sigma_i^2}} \tag{5}$$

Note that s_i only depends on σ_i/α. See also figure 2. Alternatively, σ_i can be expressed in terms of s_i and α:

$$\sigma_i = \frac{\sqrt{1 - s_i^2}}{s_i} \cdot \alpha \tag{6}$$

To obtain the optimal weights, we minimize the average loss

$$E\left[L_\alpha\left(\sum_{i=1}^{n} c_i \left(X_i - \mu\right)\right)\right] \tag{7}$$

under the boundary condition $\sum_{i=1}^{n} c_i = 1$. This becomes a straightforward problem of calculating the expectation and finding the corresponding Lagrange multipliers. The computational details can be found in the appendix. In the end, we obtain

$$c_i = \left(\sum_{j=1}^{n} \frac{\sigma_i^2}{\sigma_j^2} \right)^{-1}. \tag{8}$$

This establishes an inverse proportionality between variance and optimal relative weight. By making use of (5), we also get

$$c_i = \left(\sum_{j=1}^{n} \frac{s_j^2}{1 - s_j^2} \right)^{-1} \cdot \frac{s_i^2}{1 - s_i^2}. \tag{9}$$

Two things are worth noting. First, the scale parameter α has vanished from equation (9). That is, as long as the loss function has the *structure* given by (1), we obtain the same optimal relative weights. Arguably, this property is a substantial asset of our approach: The optimal weights do not depend on the scale parameter α that specifies the inflection point of the loss function. So even if the exact form of an appropriate loss function is disputed, our results can be applied.

Second, the weights in (8) equal the optimal weights that would have been obtained if one had used quadratic loss instead of our loss function L (see the appendix). So we obtain the surprising result that in the case under investigation, the recommendations under our realistic loss function and the recommendations under a conventional loss function agree. It is a project for further research to generalize this result, e.g. by allowing the X_i to be non-normal.

4 Conclusions

What did we achieve? We have set up a model where individual judgments, predictions or measurements are pooled into a single verdict. Such problems are pervasive in politics, economy, and science – at any place where different pieces of information have to be combined. Within our model, we have then calculated which relative weights lead to a minimal expected loss, if we know the agents' degree of expertise.

Let us stress two main points. First, we chose loss functions that are, due to their normalized character, much more suitable for problems of opinion pooling than the standard statistical measure of quadratic loss. This makes our approach more realistic than the standard approach. Second, our optimal weights are independent of the precise loss function in this family. Hence, even if there is uncertainty about the exact loss rate, our results keep their normative force. Therefore, we believe that our model is a fruitful contribution to solving problems of pooling information.

A Proofs

We assume that the random variables X_i are normally distributed with common mean ($X_i \sim N(\mu, \sigma_i)$). From equations (3) and (4), we obtain:

$$\begin{aligned}
s_i &:= E\left[S_\alpha(X_i - \mu)\right] \\
&= \frac{1}{\sqrt{2\pi}\,\sigma_i} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma_i^2}(x-\mu)^2} \cdot e^{-\frac{1}{2\alpha^2}(x-\mu)^2} dx \\
&= \frac{1}{\sqrt{2\pi}\,\sigma_i} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{1}{\sigma_i^2}+\frac{1}{\alpha^2}\right)(x-\mu)^2} dx
\end{aligned}$$

We introduce the new variable Σ_i,

$$\Sigma_i^{-1} := \sqrt{\frac{1}{\sigma_i^2} + \frac{1}{\alpha^2}}\,, \tag{10}$$

and obtain:

$$\begin{aligned}
s_i &= \frac{1}{\sqrt{2\pi}\,\sigma_i} \int_{-\infty}^{\infty} e^{-\frac{1}{2\Sigma_i^2}(x-\mu)^2} dx \\
&= \frac{\sqrt{2\pi}\,\Sigma_i}{\sqrt{2\pi}\,\sigma_i} = \frac{\Sigma_i}{\sigma_i}
\end{aligned}$$

Using equation (10), we finally obtain

$$s_i = \frac{\alpha}{\sqrt{\alpha^2 + \sigma_i^2}}. \tag{11}$$

The other equation follows by resolving this equation for σ_i.

Now, we tackle the optimizing problem, and we calculate the variance of $\sum c_i\, X_i$. It is straightforward to show (and it holds for all independent random variables) that

$$\begin{aligned}
V\left(\sum_{i=1}^{n} c_i X_i\right) &= \mathbb{E}\left[\left(\sum_{i=1}^{n} c_i\, X_i - \mu)\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{n} c_i\,(X_i - \mu)\right)^2\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i\, c_j\, \mathbb{E}\left[(X_i - \mu)(X_j - \mu)\right] \\
&= \sum_{i=1}^{n} c_i^2\, \mathbb{E}\left[(X_i - \mu)^2\right] \\
&= \sum_{i=1}^{n} c_i^2\, \sigma_i^2.
\end{aligned}$$

Thus, the random variable $\sum_{i=1}^{n} c_i\, X_i$ is distributed according to $N(\mu, \sigma^2)$ where $\sigma^2 := \sum_{i=1}^{n} c_i^2 \sigma_i^2$.

Combining this result with equation (11), we obtain

$$\begin{aligned}\mathbb{E}\left[L_\alpha\left(\sum_{i=1}^n c_i X_i - \mu\right)\right] &= 1 - \mathbb{E}\left[S_\alpha\left(\sum_{i=1}^n c_i X_i - \mu\right)\right] \\ &= 1 - \frac{\alpha}{\sqrt{\alpha^2 + \sum_{i=1}^n c_i^2\, \sigma_i^2}}. \qquad (12)\end{aligned}$$

It is well known (Lehrer and Wagner 1981, 139) that $\sum_{i=1}^n c_i^2 \sigma_i^2$ is minimized under the constraint $\sum c_i = 1$ by setting

$$c_i = \left(\sum_{j=1}^n \frac{\sigma_i^2}{\sigma_j^2}\right)^{-1}. \qquad (13)$$

This implies the desired result since (12) is monotonously increasing in $\sum_{i=1}^n c_i^2\, \sigma_i^2$. Therefore the left hand side of (12) is minimized by the expression in (13). □

References

Hoeting, Jennifer, David Madigan, Adrian Raftery, Chris Volinsky (1999): "Bayesian Model Averaging: A Tutorial". *Statistical Science* 14: 382–417.

Kahneman, Daniel, and Amos Tversky (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty", *Journal of Risk and Uncertainty* 5: 297-323.

Kahneman, Daniel, and Amos Tversky (2000): *Choices, Values and Frames*, Cambridge: Cambridge University Press.

Lehrer, Keith, and Carl Wagner (1981): *Rational Consensus in Science and Society.* Reidel: Dordrecht.

Shapley, Lloyd, and Bernard Grofman (1984): "Optimizing Group Judgmental Accuracy in the Presence of Interdependence", *Public Choice* 43: 329–343.