A Loss Function
for Box-Constrained Inverses Problems

Kiyoshi Yoneda*

Abstract. A loss function is proposed for solving box-constrained inverse problems. Given causality mechanisms between inputs and outputs as smooth functions, an inverse problem demands to adjust the input levels to make the output levels as close as possible to the target values; box-constrained refers to the requirement that all outcome levels remain within their respective permissible intervals. A feasible solution is assumed known, which is often the *status quo*. We propose a loss function which avoids activation of the constraints. A practical advantage of this approach over the usual weighted least squares is that permissible outcome intervals are required in place of target importance weights, facilitating data acquisition. The proposed loss function is smooth and strictly convex with closed-form gradient and Hessian, permitting Newton family algorithms. The author has not been able to locate in the literature the Gibbs distribution corresponding to the loss function. The loss function is closely related to the *generalized matching law* in psychology.

1. INTRODUCTION

An *inverse problem* demands that given a *causality* mechanism and a number of resulting *effects* their *causes* be identified. The causality is modeled as a finite dimensional input-output system $\mathbb{R}^m \ni \mathbf{x} \overset{\mathbf{f}}{\mapsto} \mathbf{y} \in \mathbb{R}^n$ or, writing for each output item, $f_i : \mathbb{R}^m \ni \mathbf{x} \mapsto y_i \in \mathbb{R}, \ 1 \le i \le n$: The vector of causes $\mathbf{x}$ goes through causality mechanisms f_i and results in the effects $f_i(\mathbf{x})$; although the causes always precede the effects in time the model does not express this feature explicitly. Throughout this paper the vectors are in bold face letters and their elements are in corresponding normal letters. The input itself is to be considered a part of the output: "If I decide to do something, everything that follows is a consequence, including the action itself" (Layard, 2006) p. 119, viz., $f_j : x_j \mapsto y_j = x_j, \ 1 \le j \le m$ implying $m \le n$. The

* Fukuoka University Faculty of Economics 8-19-1 Nanakuma, Jounan-ku, Fukuoka, 814-0180 Japan. E-mail: yoneda@econ.fukuoka-u.ac.jp

present input and output values, $\mathbf{x}^0$ and $\mathbf{y}^0$, and the mechanism $\mathbf{f}()$ which links them are assumed known.

Suppose the circumstances have changed so that the desired output values are now $\mathbf{y}^1$, which represents small deviations from the present values $\mathbf{y}^0$ in the sense that the linear approximation:

$$\mathbf{y}^1 \approx \mathbf{f}(\mathbf{x}^0) + J_{\mathbf{f}}(\mathbf{x}^0)(\mathbf{x}^1 - \mathbf{x}^0) \qquad\qquad 1 \le i \le n$$

holds, where:

$$
J_{\mathbf{f}}(\mathbf{x}^0) := \begin{bmatrix} \vdots \\ \nabla \mathrm{f}_i(\mathbf{x}^0) \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 & & & 0 \\ & \ddots & & \\ 0 & & & 1 \\ & & \vdots & \\ \cdots & \frac{\partial \mathrm{f}_i}{\partial \mathrm{x}_j}(\mathbf{x}^0) & \cdots \\ & & \vdots & \end{bmatrix}
$$

is the Jacobian matrix of $\mathbf{f}$ at $\mathbf{x}^0$; "$\approx$" indicates near-equality. The top identity part is for $\mathrm{f}_j : x_j \mapsto x_j$. The decision maker controls n aspects of life by adjusting its subset of size m $(\le n)$, with no such rigid budget constraint as found in microeconomics model of individual behavior, e.g. (Wichers, 1996).

Now the problem is to find an $\mathbf{x}$ such that $J\mathbf{x} \approx \mathbf{y}^1 - (\mathbf{y}^0 - J\mathbf{x}^0)$ given $J := J_{\mathbf{f}}(\mathbf{x}^0)$ and $\mathbf{y}^1 - (\mathbf{y}^0 - J\mathbf{x}^0)$. This is an ordinary *linear inverse problem* aside from the top part of J being a unit matrix. The usual method of solution would be the least squares with special considerations to the stability of the solution, i.e., that a small change in $\mathbf{y}^1$ should not incur a huge change in the solution $\mathbf{x}$.

A problem with using the least squares method to solve such inverse problems is that the solution often lies beyond the practical limits. In practice the acceptable input values $\mathbf{x}$ and corresponding output values $\mathbf{y}^0 + J(\mathbf{x} - \mathbf{x}^0)$ are bounded, more often than not: one cannot, for instance, purchase a negative amount of product or sleep more than 24 hours a day. That kind of knowledge may be incorporated into a formulation by requiring that the output values remain within permissible intervals:

Definition 1. Box-constrained inverse problem
Given J, $\mathbf{x}^0$, $\mathbf{y}^0$, and $\mathbf{y}^1$ find an $\mathbf{x}$ such that:

$$\left(\mathbf{y}^0 - J\mathbf{x}^0\right) + J\mathbf{x} \approx \mathbf{y}^1 \tag{1}$$

satisfying the *box-constraints*:

$$\mathbf{a} < \left(\mathbf{y}^0 - J\mathbf{x}^0\right) + J\mathbf{x} < \mathbf{b} \tag{2}$$

Finding a feasible solution is trivial when the *status quo* is within the permissible range and that only the target values $\mathbf{y}$ change from $\mathbf{y}^0$ to $\mathbf{y}^1$ within the box constraints, which is often the case, but can be difficult otherwise.

Two successful lines of attack to such problems are generalized linear models (McCullagh and Nelder, 1989) and maximum entropy models (Campbell and Hill, 2006), (Golan et al., 1996). They both are conceptually clear but not quite computationally efficient since the former reduces to iterated least squares and the latter requires two or more support variables to represent a single parameter x_j in the equations.

This paper proposes a method which is both conceptually clear and computationally more efficient. The approach consists of designing a smooth strictly convex loss function such that attains a minimum of zero when all targets are met and diverges at the box constraints avoiding them to become active.

The remainder of this paper is organized as follows. To simplify notation Section 2 defines a standard form for the box-constrained inverse problems. A minimal numerical example is introduced. The near-equality is specified in Section 3 by defining a loss function. Section 4 discusses how to convert the target positions relative to their bounds into the importance weight parameters in the loss function. Section 5 points out some properties of the loss function. The paper is summarized in Section 6. Appendix A calculates the gradient and the Hessian of the proposed loss function. Appendix B attempts to endow a geometric interpretation to the loss function.

2. THE STANDARD FORM

This section puts the problem into a *standard form* involving no constants a_i and b_i in order to simplify notation. We use dot subscripts for writing matrices in row and column vectors:

$$J = \begin{bmatrix} & \vdots & \\ \cdots & J_{ij} & \cdots \\ & \vdots & \end{bmatrix} = \begin{bmatrix} \vdots \\ \mathbf{J}_{i\cdot} \\ \vdots \end{bmatrix} = \begin{bmatrix} \cdots & \mathbf{J}_{\cdot j} & \cdots \end{bmatrix}.$$

Since $(y_i^0 - \mathbf{J}_{i\cdot} \mathbf{x}^0) + \mathbf{J}_{i\cdot} \mathbf{x} \approx y_i^1$ and $a_i < y_i^1 < b_i$,

$$\frac{(y_i^0 - \mathbf{J}_{i\cdot} \mathbf{x}^0) - a_i + \mathbf{J}_{i\cdot} \mathbf{x}}{b_i - a_i} \approx \frac{y_i^1 - a_i}{b_i - a_i}.$$

The right hand side of the near-equation fits into the open unit interval $]0, 1[$; the left hand side must also fit into the same open unit interval. By letting:

$$z_i := \frac{y_i^1 - a_i}{b_i - a_i} \qquad\qquad 0 < z_i < 1 \qquad\qquad 1 \le i \le n$$

$$\theta_j := \frac{x_j - a_j}{b_j - a_j} \qquad\qquad 0 < \theta_j < 1 \qquad\qquad 1 \le j \le m \ ,$$

$$\frac{y_i^0 - \mathbf{J}_{i\cdot} \mathbf{x}^0 - a_i + \sum_j J_{ij} a_j + \sum_j J_{ij} (b_j - a_j)\theta_j}{b_i - a_i} \approx z_i.$$

The box-constrained inverse problem (1) with (2) has been reduced to:

Definition 2. The standard form box-constrained linear inverse problem

$$\mathbf{c} + D\,\boldsymbol{\theta} \approx \mathbf{z} \quad \text{or} \quad \zeta_i := \hat{z}_i(\boldsymbol{\theta}) := c_i + D_{i\,.}\,\boldsymbol{\theta} \approx z_i$$

$$\zeta_i, \hat{z}_i(\boldsymbol{\theta}), z_i, \theta_j \in]0,1[\quad 1 \le i \le n \quad 1 \le j \le m \le n.$$

$\square$

Here:

$$\mathbf{c} := L \begin{bmatrix} \vdots \\ y_i^0 - \mathbf{J}_{i\,.}\,\mathbf{x}^0 + \sum_{j=1}^m J_{ij}a_j - a_i \\ \vdots \end{bmatrix} \qquad D := L\,J\,R \qquad \mathbf{z} = L(\mathbf{y} - \mathbf{a})$$

$$L := \begin{bmatrix} b_1 - a_1 & & \\ & \ddots & \\ & & b_n - a_n \end{bmatrix}^{-1} \qquad R := \begin{bmatrix} b_1 - a_1 & & \\ & \ddots & \\ & & b_m - a_m \end{bmatrix}$$

and we assume to have a feasible solution $\boldsymbol{\theta}^0$ available, where $\theta_j^0 := \frac{x_j^0 - a_j}{b_j - a_j} \in]0,1[$ for $1 \le j \le m$. Once the solution $\hat{\boldsymbol{\theta}}$ is found, the solution to the original problem (1) with (2) may be secured by:

$$\hat{\mathbf{x}} := \mathbf{a}_m + R\,\hat{\boldsymbol{\theta}} \,, \qquad\qquad \mathbf{a}_m := [a_1 \ \cdots \ a_m]' \qquad\qquad (3)$$

provided that the transformation preserves the optimality of the solution.

The rest of this section comprises a numerical example to be used throughout this paper.

2.1. A MINIMAL LINEAR INVERSE PROBLEM

In your company a quantity x_1 of ingredient 1 and a quantity x_2 of ingredient 2 are processed into a quantity $2x_1 + x_2$ of product. The purchase department is going to place orders of the ingredients for the coming period. You normally place an order of a unit of each ingredients to manufacture three units of product. However, the sale predicted for this coming period is low, with only one unit. To stabilize the relationship with the ingredient suppliers the purchase department wishes to order the same amount of ingredients as usual, whereas the sales department demands that the production be reduced to one unit.

The usual purchase is $\mathbf{x}^0 = [1 \ 1]'$ and $\mathbf{y}^0 = J\mathbf{x}^0 = [1 \ 1 \ 3]'$, where the prime denotes transposition. Now $y_1^1 = y_2^1 = 1$ by the purchase department but the sales department says $y_3^1 = 1$. So the system of approximate equations to solve is

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \approx \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

The direction of causality is unspecified among the near-equations, whether the cost of two ingredients sum up to the total y_3^1 or the budget y_3^1 is to be allocated to both ingredients.

2.2. A WEIGHTED LEAST SQUARES SOLUTION

The first two equations are approximate since your order can differ from the usual. The last equation is also approximate since the prediction may be inaccurate and stock may absorb unsold or excessive product.

Suppose your purchase department considers a $\pm 1/5$ fluctuation from y_1 to be normal. The tolerance is higher for y_2, which is $\pm 1/2$. As with the production you want to allow a $\pm 1/10$ standard error. Treating these values as standard deviations $\sigma := [1/5\ 1/2\ 1/10]'$, the usual weights for the weighted least squares are $[25\ 4\ 100]'$. The weighted least squares solution for this problem is $\hat{\mathbf{x}} = [0.62\ -0.19]'$: even though all numbers appearing in the problem are nonnegative, the amount of ingredient 2 to purchase is negative. The common practice of coercing negative values into zero, like $x_2 := 0$, causes difficulties: the balance of quantity between ingredients 1 and 2 will be disrupted; the purchase department will worry about the relationship with the suppliers of ingredient 2 which would be cut altogether.

2.3. A BOX-CONSTRAINED FORMULATION

The purchase department considers that ordering 0.20 units or less of ingredient 1 will be harmful to the relationship with the suppliers, and that ordering twice the usual amount or more would also be unrealistic. Ingredient 2 should be ordered a positive amount, but not two or more units. The sales department accepts that the production be between 0.70 and 1.30 units. You wish to place an order of the ingredients so that:

$$
\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \approx \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} , \qquad \begin{bmatrix} 0.20 \\ 0.00 \\ 0.70 \end{bmatrix} < \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} < \begin{bmatrix} 2.00 \\ 2.00 \\ 1.30 \end{bmatrix} .
$$

The usual amount $\mathbf{x}^0 = [1\ 1]'$ does not satisfy the box constraints, whereas reducing it to about $1/3$, $\mathbf{x}^0 = [0.3\ 0.3]'$ does.

2.4. IN STANDARD FORM

Since:

$$
L = \begin{bmatrix} 0.556 & 0 & 0 \\ 0 & 0.500 & 0 \\ 0 & 0 & 1.667 \end{bmatrix} \qquad J = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} \qquad R = \begin{bmatrix} 1.8 & 0 \\ 0 & 2 \end{bmatrix} ,
$$

$$
D = LJR = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 6 & 3.333 \end{bmatrix} \qquad \mathbf{c} = - \begin{bmatrix} 0 \\ 0 \\ 0.3 \end{bmatrix} \qquad \mathbf{z} = \begin{bmatrix} .444 \\ .500 \\ .500 \end{bmatrix}
$$

the problem in the standard form is:

$$
- \begin{bmatrix} 0 \\ 0 \\ 0.3 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 6 & 3.333 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \approx \begin{bmatrix} .444 \\ .500 \\ .500 \end{bmatrix} ,
$$

with the initial value $\boldsymbol{\theta}^0 = R^{-1}(\mathbf{x}^0 - \mathbf{a}_m) = [.056 \ .15]'$ which is feasible: $\mathbf{c} + D\boldsymbol{\theta}^0 = [.056 \ .15 \ .33]' \in \,]0,1[^3$.

3. THE LOSS FUNCTION

Assume an additive loss function and define the solution by:

$$\hat{\boldsymbol{\theta}} := \arg \min_{\boldsymbol{\theta} \in]0,1[^m} \sum \mathrm{d}(\zeta_i, z_i) \qquad \zeta_i := \hat{z}_i(\boldsymbol{\theta}) := c_i + D_i.\boldsymbol{\theta} \qquad 0 \le \mathrm{d}(\cdot, \cdot).$$

For the method of least squares $\mathrm{d}(\zeta_i, z_i) := (\zeta_i - z_i)^2$, but this is inadequate for our case since $\boldsymbol{\theta} \in]0,1[^m$ may not be met, as has been illustrated by the example in the previous section. One could set up brute force constraints, but that would imply extra computational complexity in addition to the necessity of a convincing interpretation of the method: e.g., what does it mean to coerce a quantity into a nonnegative value when the objective function does permit a negative argument? Our task now is to find a reasonable *individual loss function* $\mathrm{d}(\zeta_i, z_i)$ which measures how far $\zeta_i \in]0,1[$ is from $z_i \in]0,1[$, such that avoids activation of the constraints. This section derives the individual loss function $\mathrm{d}(\cdot, \cdot)$ in an intuitive way.

The conventional individual loss functions used to satisfy $0 \le \zeta_j$, but not necessarily $\zeta_j < 1$, is the χ^2 defined by $\mathrm{d}(\zeta_i, z_i) = \left(\frac{\zeta_i - z_i}{z_i}\right)^2 = (r_i - 1)^2$ where $r_i := \frac{\zeta_i}{z_i}$. For strict positivity $0 < \zeta_j$ the log-square $\mathrm{d}(\zeta_i, z_i) = (\log \zeta_i - \log z_i)^2 = (\log r_i)^2$ is popular. Their hybrid:

$$\mathrm{d}(\zeta_i, z_i) = \left(\frac{\zeta_i - z_i}{z_i}\right)(\log \zeta_i - \log z_i) = (r_i - 1)\log r_i \qquad (4)$$

called the *semilog loss* has been proposed in (Yoneda, 2006). The three are illustrated in Figure 1.

Since the logarithmic function is defined only for the positive reals, $0 < r_i$ must be satisfied in the rightmost expression in (4). The semilog loss (4) is asymmetric, i.e., $\mathrm{d}(\hat{z}_i, z_i)$ generally differs from $\mathrm{d}(z_i, \hat{z}_i)$. The intuitive meaning of this function is as follows. While $r_i - 1 \approx 0$ states that ζ_i is close to z_i in terms of the Euclidean distance measured by z_i as the unit, $\log r_i \approx 0$ states that ζ_i is close to z_i in terms of ratio of ζ_i against z_i. The hybrid $(r_i - 1)\log r_i \approx 0$ states that ζ_i is close to z_i in both senses. Dropping the subscripts, schematically:

$$r \approx 1 \quad \Leftrightarrow \quad [\,r - 1 \approx 0 \quad \text{and} \quad \log r \approx 0\,] \quad \Rightarrow \quad (r - 1)\log r \approx 0,$$

since all functions involved are smooth. Note that since $r - 1$ and $\log r$ have the same sign, $0 \le (r - 1)\log r$.

The graph for $(r - 1)\log r = \left(\frac{\zeta}{z} - 1\right)\log \frac{\zeta}{z}$, the solid line in Figure 1, appears as the dashed line in Figure 2 for $z = \frac{1}{3}$. This satisfies $0 < \zeta$ but not $\zeta < 1$.
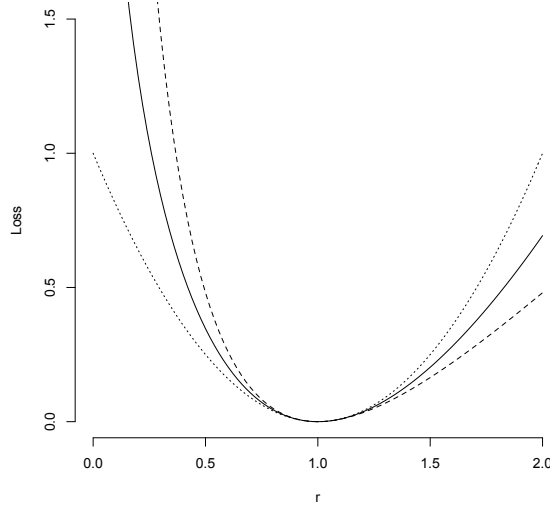
Fig. 1. Individual loss function $(r_i - 1) \log r_i$.
The semilog loss (4) (solid), log-squared (dashed), and χ^2 (dotted)

By flipping the dashed graph in Figure 2 horizontally, then shifting it to the right by one, and adjusting the minimum to the same z, another individual loss function $\left(\frac{1-\zeta}{1-z} - 1\right) \log \frac{1-\zeta}{1-z}$ may be constructed, illustrated as the dotted line in Figure 2. This satisfies $\zeta < 1$ but not $0 < \zeta$. By adding the two, a new individual loss function:

$$\mathrm{d}(\zeta, z) := \left(\frac{\zeta}{z} - 1\right) \log \frac{\zeta}{z} + \left(\frac{1-\zeta}{1-z} - 1\right) \log \frac{1-\zeta}{1-z} \tag{5}$$

is defined, illustrated as the solid line in Figure 2, which is the sum of the dashed and the dotted lines. This satisfies $0 < \zeta < 1$. Geometric interpretations of the semilog (4) and the proposed loss functions are described in Appendix B.

Proposition 1. *The individual loss function* (5) *is strictly convex in* $\zeta \in]0, 1[$.

Proof. Since $\left(\frac{\zeta}{z} - 1\right) \log \frac{\zeta}{z}$ is strictly convex in $\zeta \in]0, \infty[$ and $\left(\frac{1-\zeta}{1-z} - 1\right) \log \frac{1-\zeta}{1-z}$ is strictly convex in $\zeta \in]-\infty, 1[$, their sum (5) is strictly convex in $\zeta \in]0, 1[$. $\qquad\square$

Figure 3 illustrates $w \, \mathrm{d}(\zeta, z)$, where d is defined by (5), for different values of w: a larger w corresponds to a steeper loss function; the parameter w represents the importance of the target. Since the permissible interval for outcome ζ has been standardized, the weight w depends solely on the location of the target z within the unit interval $]0, 1[$. The problem of converting the target position z into the importance weight w will be considered in the next section.
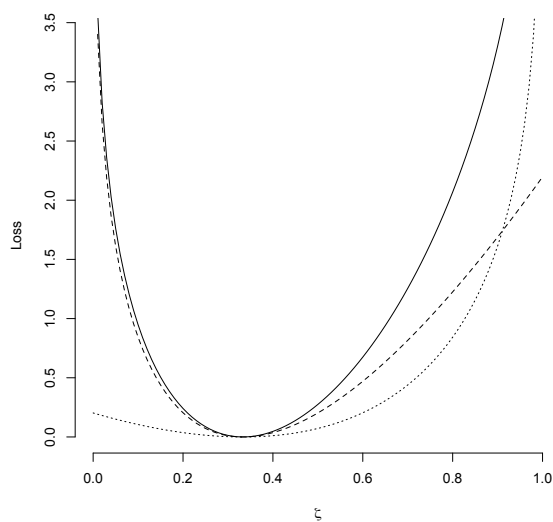
Fig. 2. Loss function over the open unit interval.
The loss (5) (solid), its first term (dashed), and the second term (dotted),
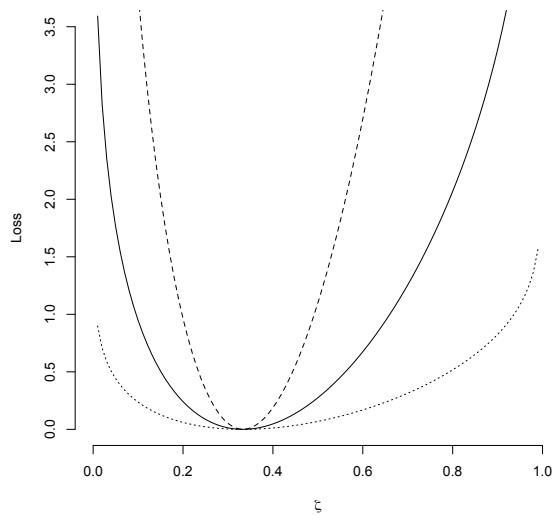all with $z = \frac{1}{3} \in {]}0, 1{[}$

Fig. 3. $w\,\mathrm{d}(\zeta, \frac{1}{3})$ (dotdash) $w = 4$ (dashed), $= 1$ (solid), $= \frac{1}{4}$ (dotted).
The lines correspond to the distributions in Figure 4

Definition 3. A solution $\hat{\boldsymbol{\theta}}$ to the standard form box-constrained linear inverse problem is defined by:

$$\hat{\boldsymbol{\theta}} := \arg \min_{\boldsymbol{\theta} \in]0,1[^m} \ell(\boldsymbol{\theta}) , \quad \ell(\boldsymbol{\theta}) := \sum \ell_i(\boldsymbol{\theta}) \tag{6}$$

$$\ell_i(\boldsymbol{\theta}) := w_i \left\{ \left(\frac{\zeta_i}{z_i} - 1 \right) \log \frac{\zeta_i}{z_i} + \left(\frac{1 - \zeta_i}{1 - z_i} - 1 \right) \log \frac{1 - \zeta_i}{1 - z_i} \right\} \tag{7}$$

$$\zeta_i := c_i + D_{i.}\, \boldsymbol{\theta}$$

where:

$$\zeta_i, z_i, \theta_j \in]0,1[\qquad\qquad 1 \le i \le n \qquad\qquad 1 \le j \le m \le n.$$

$\square$

Proposition 2. *If a feasible solution exists, then $\hat{\boldsymbol{\theta}}$ exists uniquely.* $\square$

This is obvious since the loss function is strictly convex. Now it remains to check that (3) provides a solution to the original problem:

Proposition 3. *The standard form solution $\hat{\boldsymbol{\theta}}$ together with (3) uniquely determines a solution $\hat{\mathbf{x}}$ to the original box-constrained inverse problem such that minimizes the loss function.*

Proof. Since (3) is a non-degenerate affine transformation, the loss function $\ell(\boldsymbol{\theta}(\mathbf{x}))$ in $\mathbf{x}$ is strictly convex. Now $\frac{\partial \ell(\boldsymbol{\theta}(\mathbf{x}))}{\partial \mathbf{x}_k} = \sum_i \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \theta_j} \frac{\partial \theta_j}{\partial x_k} = \sum_i \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \theta_j} R_{jk}^{-1}$, it follows that $\left[\forall j \; \frac{\partial \ell(\boldsymbol{\theta}(\mathbf{x}))}{\partial \mathbf{x}_j} = 0 \right] \Leftrightarrow \left[\forall j \; \frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_j} = 0 \right]$. $\square$

Proposition 4. *The Gibbs distribution corresponding to the individual loss function (7) is:*

$$\varphi(\zeta) \propto \left(\frac{\left(\frac{\zeta}{z} \right)^{\frac{1}{z}}}{\left(\frac{1-\zeta}{1-z} \right)^{\frac{1}{1-z}}} \right)^{w(z-\zeta)} \qquad\qquad \int_0^1 \varphi(\zeta) d\zeta = 1. \tag{8}$$

Proof.

$$\varphi(\zeta) \propto e^{-w\, \mathrm{d}(\zeta,z)} = e^{-w \left\{ \left(\frac{\zeta}{z} - 1 \right) \log \frac{\zeta}{z} + \left(\frac{1-\zeta}{1-z} - 1 \right) \log \frac{1-\zeta}{1-z} \right\}} = \left\{ \left(\frac{\zeta}{z} \right)^{\frac{1}{z}} \left(\frac{1-\zeta}{1-z} \right)^{-\frac{1}{1-z}} \right\}^{w(z-\zeta)}$$

$\square$

The resulting distribution (8) is similar to but differs from the beta distribution. Figure 4 illustrates how the weight w affects the shape of (8).
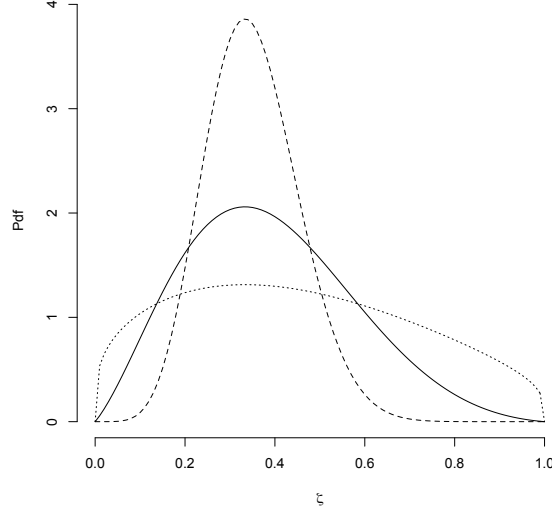
Fig. 4. Pdf $\propto e^{-w\mathrm{d}(\zeta,z)}$, $z = \frac{1}{3}$ (dotdash) $w = 4$ (dashed), $= 1$ (solid), $= \frac{1}{4}$ (dotted).
The lines correspond to the loss functions in Figure 3

Remark.

$$\varphi(\zeta; z, w) \xrightarrow{w \longrightarrow 0} \textit{Uniform distribution in }]0, 1[$$

$$\varphi(\zeta; z, w) \xrightarrow{w \longrightarrow \infty} \delta(\zeta - z); \ \textit{Dirac's delta.}$$

$\square$

Since the Gibbs distribution $\propto e^{-\ell(\boldsymbol{\theta})}$ is the maximum entropy distribution satisfying $\mathrm{E}[\ell(\boldsymbol{\theta})] = \mathrm{constant}$, a reasonable interpretation of the solution $\hat{\mathbf{x}}$ would be that the decision maker chooses those parameter values with highest probability under the same circumstances but the actual values chosen will vary, rather than that the individual invariably chooses the prescribed parameter values.

4. TARGET-TO-WEIGHT CONVERSION

In order to fix a solution of (6) we need to determine the weights $\mathbf{w} := [\cdots \ w_i \ \cdots]$ as has been pointed out in the previous section. The triangular distribution:

$$\tau(\zeta) := \begin{cases} 2\frac{\zeta}{z} & 0 \leq \zeta \leq z \\ 2\frac{1-\zeta}{1-z} & z \leq \zeta \leq 1, \end{cases} \tag{9}$$

illustrated by the dotted line in Figure 6 is arguably the simplest model of a continuous pdf with a peak at z and fades to zero at both ends 0 and 1. Hence its corresponding loss function:

$$
\ell_\tau(\zeta) := \begin{cases} -\log \frac{\zeta}{z} & 0 < \zeta \le z \\ -\log \frac{1-\zeta}{1-z} & z \le \zeta < 1 \end{cases}
$$

is a natural candidate to adopt instead of (5).

There are two reasons why we do not simply do so. The first is computational convenience. A smooth strictly convex loss function permits highly efficient Newton family of minimization methods while the above loss function requires subgradient methods for minimization, impeding solution of large problems. The second is that results in psychology called the *modern matching law* (Prelec, 1984), which is a generalization of the classic *matching law* (Herrnstein et al., 1997), seem to support the proposed pdf better than the triangular pdf. For these reasons we approximate the triangular distribution $\tau()$ in (9), illustrated by the dotted line in Figure 6, by the smooth Gibbs distribution $\varphi()$ in (8), , illustrated in Figure 4 and as the solid line in Figure 6.

The weight w is set so that the distribution corresponding to the loss function best approximates the triangular distribution in the sense of Kullback-Leibler information:

$$
K(\tau, \varphi) := E_\tau\left[\log \frac{\tau(\tilde{\zeta})}{\varphi(\tilde{\zeta})}\right] = \int_{-\infty}^{\infty} \tau(\zeta) \log \frac{\tau(\zeta)}{\varphi(\zeta)} d\zeta.
$$

Note that $K(\tau, \varphi)$ may be negative.

Proposition 5. *For Gibbs distributions $\tau(\zeta) = e^{-\ell_\tau(\zeta)}/Z_\tau$ and $\varphi(\zeta) = e^{-\ell_\varphi(\zeta)}/Z_\varphi$,*

$$
E_\tau[\ell_\varphi(\tilde{\zeta}) - \ell_\tau(\tilde{\zeta})] = K(\tau, \varphi) + \log \frac{Z_\tau}{Z_\varphi} \tag{10}
$$

where $\tilde{\zeta}$ is a random variable; Z_τ and Z_φ are normalization constants.

Proof.

$$
\begin{aligned}
K(\tau, \varphi) &= \int_{-\infty}^{\infty} \{\ell_\varphi(\zeta) + \log Z_\varphi - \ell_\tau(\zeta) - \log Z_\tau\}\tau(\zeta)d\zeta \\
&= E_\tau[\ell_\varphi(\tilde{\zeta}) - \ell_\tau(\tilde{\zeta})] + \log(Z_\varphi/Z_\tau).
\end{aligned}
$$

$\square$

This implies that the following are equivalent: to approximate a loss function by another by minimizing the expected difference between the two, and to approximate the distribution corresponding to a loss function by the distribution corresponding to another loss function by information minimization. The triangular distribution τ may now be approximated by the distribution φ corresponding to the loss function $\ell_\varphi := \ell$ by requiring the best fit:

Proposition 6.

$$\mathrm{E}_\tau\left[\ell_\varphi(\tilde\zeta)\right] = \mathrm{E}_\tau\left[\ell_\tau(\tilde\zeta)\right]$$

$$\Rightarrow \quad w = \mathrm{w}(z) := \frac{9z(1-z)}{6\{(3z-1)\log z - (3z-2)\log(1-z)\} + 44z(1-z) - 5} \quad (11)$$

Proof. Since:

$$\mathrm{E}_\tau\left[\ell_\tau(\tilde\zeta)\right] = 2\left(-\int_0^z \frac{\zeta}{z}\log\frac{\zeta}{z}\,d\zeta + \int_z^1 \frac{1-\zeta}{1-z}\log\frac{1-\zeta}{1-z}\,d\zeta\right)$$

$$= 2\left(\frac{z}{4} - \frac{1-z}{4}\right) = \frac{z-(z-1)}{2} = \frac{1}{2},$$

by:

$$\mathrm{E}_\tau\left[\ell_\varphi(\tilde\zeta)\right] = \mathrm{E}_\tau\left[\ell_\tau(\tilde\zeta)\right],$$

$$\mathrm{E}_\tau\left[\ell_\varphi(\tilde\zeta)\right] = w\,\mathrm{E}_\tau\left[\left(\frac{\zeta}{z}-1\right)\log\frac{\zeta}{z} + \left(\frac{1-\zeta}{1-z}-1\right)\log\frac{1-\zeta}{1-z}\right] = \frac{1}{2} \quad (12)$$

Now $\int t^n \log t\,dt = \frac{t^{n+1}}{n+1}\left(\log t - \frac{1}{n+1}\right) + \text{constant}$ for $0 \le n$, so that:

$$\mathrm{E}_\tau\left[\left(\frac{\zeta}{z}-1\right)\log\frac{\zeta}{z} + \left(\frac{1-\zeta}{1-z}-1\right)\log\frac{1-\zeta}{1-z}\right] =$$

$$= \int_0^1 \tau(\zeta)\left\{\left(\frac{\zeta}{z}-1\right)\log\frac{\zeta}{z} + \left(\frac{1-\zeta}{1-z}-1\right)\log\frac{1-\zeta}{1-z}\right\}d\zeta =$$

$$= \quad \int_0^1 \tau(\zeta)\frac{\zeta}{z}\log\frac{\zeta}{z}d\zeta \qquad - \int_0^1 \tau(\zeta)\log\frac{\zeta}{z}d\zeta$$

$$+ \int_0^1 \tau(\zeta)\frac{1-\zeta}{1-z}\log\frac{1-\zeta}{1-z}d\zeta - \int_0^1 \tau(\zeta)\log\frac{1-\zeta}{1-z}d\zeta =$$

$$= 2\left[\quad \int_0^z \left(\frac{\zeta}{z}\right)^2\log\frac{\zeta}{z}d\zeta \qquad + \int_z^1 \frac{1-z}{1-z}\cdot\frac{\zeta}{z}\log\frac{\zeta}{z}d\zeta\right.$$

$$- \int_0^z \frac{\zeta}{z}\log\frac{\zeta}{z}d\zeta \qquad - \int_z^1 \frac{1-\zeta}{1-z}\log\frac{\zeta}{z}d\zeta$$

$$+ \int_0^z \frac{\zeta}{z}\cdot\frac{1-\zeta}{1-z}\log\frac{1-\zeta}{1-z}d\zeta + \int_z^1 \left(\frac{1-\zeta}{1-z}\right)^2\log\frac{1-\zeta}{1-z}d\zeta$$

$$\left. - \int_0^z \frac{\zeta}{z}\log\frac{1-\zeta}{1-z}d\zeta \qquad - \int_z^1 \frac{1-\zeta}{1-z}\log\frac{1-\zeta}{1-z}d\zeta \qquad\right] =$$

$$= \frac{6\{(3z-1)\log z - (3z-2)\log(1-z)\} + 44z(1-z) - 5}{18z(1-z)} \quad (13)$$

Substituting (13) into (12) yields (11). $\qquad\square$

The shape w(z) takes is as in Figure 5. The figure seems to state: if the desired value z is in the middle of the permissible range $]0, 1[$ it is important where ζ will fall; otherwise it doesn't matter much so long as ζ falls within the range. The resulting distribution for $z = 1/3$ with w given by (11) is illustrated in Figure 6.
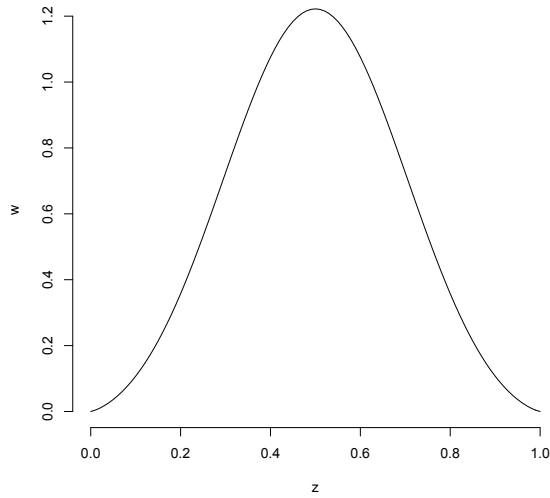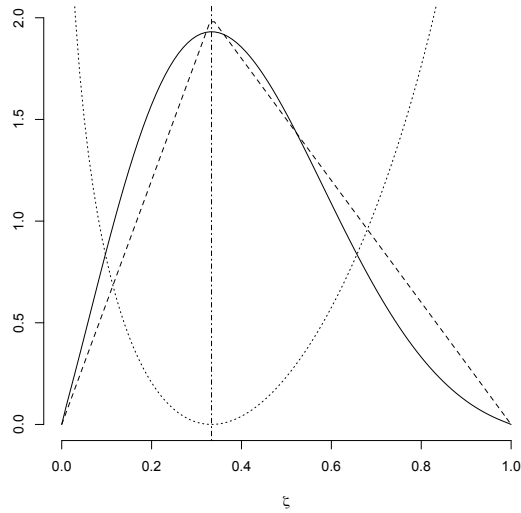

Fig. 5. Weight as a function of the mode

Fig. 6. Bounded semilog distribution $\varphi(\zeta)$ (solid), $\tau(\zeta)$ (dashed), ℓ_φ (dotted)
$z = \frac{1}{3} \in]0, 1[$ (dotdash), $w \approx 0.853$

EXAMPLE (CONTINUED FROM 2.4)

From $\mathbf{z} = [0.444\ 0.500\ 0.500]'$ and (11), $\mathbf{w} = [1.17\ 1.22\ 1.22]$.

5. PROPERTIES OF THE LOSS FUNCTION

Proposition 7. *The gradient* $\mathbf{g}$ *and the Hessian H of* $\ell(\boldsymbol{\theta})$ *are:*

$$\mathbf{g} = D'W \left[\cdots \ \tfrac{1}{z_i}\left(1 - \tfrac{1}{r_{0i}} + \log r_{0i}\right) - \tfrac{1}{1-z_i}\left(1 - \tfrac{1}{r_{1i}} + \log r_{1i}\right) \ \cdots \right]'$$
$$H = D'W \operatorname{diag}\left[\cdots \ \tfrac{1+r_{0i}}{\zeta_i^2} + \tfrac{1+r_{1i}}{(1-\zeta_i)^2} \ \cdots \right] D,$$

where:

$$\ell(\boldsymbol{\theta}) := \sum_i \ell_i(\boldsymbol{\theta}) \quad \ell_i(\boldsymbol{\theta}) := w_i\{(r_{0i} - 1)\log r_{0i} + (r_{1i} - 1)\log r_{1i}\}$$

$$r_{0i} := \frac{\zeta_i}{z_i} \quad r_{1i} := \frac{1 - \zeta_i}{1 - z_i} \quad \zeta_i := c_i + D_{i.}\boldsymbol{\theta} \quad W := \operatorname{diag}\left[\cdots \ w_i \ \cdots \right]$$

Proof. is by straightforward calculation as shown in Appendix A.   □

This permits efficient solution of the problem, including the Newton family of algorithms.

EXAMPLE (CONTINUED FROM 2.4)

From (7),

$$\hat{\boldsymbol{\theta}} := \arg \min_{\boldsymbol{\theta} \in]0,1[^2} \{\ell_1(\boldsymbol{\theta}) + \ell_2(\boldsymbol{\theta}) + \ell_3(\boldsymbol{\theta})\}$$

$$\ell_1(\boldsymbol{\theta}) := 1.17 \left\{ \left(\frac{\theta_1}{0.444} - 1\right) \log \frac{\theta_1}{0.444} + \left(\frac{1 - \theta_1}{1 - 0.444} - 1\right) \log \frac{1 - \theta_1}{1 - 0.444} \right\}$$

$$\ell_2(\boldsymbol{\theta}) := 1.22 \left\{ \left(\frac{\theta_2}{0.500} - 1\right) \log \frac{\theta_2}{0.500} + \left(\frac{1 - \theta_2}{1 - 0.500} - 1\right) \log \frac{1 - \theta_2}{1 - 0.500} \right\}$$

$$\ell_3(\boldsymbol{\theta}) := 1.22 \left\{ \left(\frac{-0.3 + (2\theta_1 + \theta_2)}{0.500} - 1\right) \log \frac{-0.3 + (2\theta_1 + \theta_2)}{0.500} + \right.$$
$$\left. \left(\frac{1 - \{-0.3 + (2\theta_1 + \theta_2)\}}{1 - 0.500} - 1\right) \log \frac{1 - \{-0.3 + (2\theta_1 + \theta_2)\}}{1 - 0.500} \right\}.$$

The solution is $\hat{\boldsymbol{\theta}} = [0.090\ 0.178]'$ with $\ell(\hat{\boldsymbol{\theta}}) = 3.25$. The corresponding solution to the original problem is $\hat{\mathbf{y}} := (\mathbf{y}^0 - J\mathbf{x}^0) + J\hat{\mathbf{x}} = (\mathbf{y}^0 - J\mathbf{x}^0) + J(\mathbf{a}_m + R\hat{\boldsymbol{\theta}}) = [0.362\ 0.355\ 1.080]'$, or buy 0.362 and 0.355 units of ingredients 1 and 2 respectively to make 1.080 units of product.

Given that usually a unit of both ingredients 1 and 2 are purchased and the production should be reduced to a third of the normal quantity, this solution meets the intuition. The interpretation is easy because all values fall within the reasonable ranges; no such thing as a negative quantity.

6. CONCLUSION

The solution method for positive inverse problems in (Yoneda, 2006) has been extended to box-constrained inverse problems, which provides a solution within a prescribed hyperbox. The method is computationally more efficient than the generalized linear models and the maximum entropy models since it requires only one variable to represent a parameter and permits Newton family of optimization methods. An important advantage of this method over the weighted least squares is that data acquisition is considerably easier: while importance weight and standard deviation are not concepts that everybody understands, almost anybody can come up with a permissible interval.

The author has not been able to locate the probability distribution (8) in the literature. Many problems remain, including the psychological interpretation of the loss function and the problem of finding a feasible solution when the *status quo* is unavailable.

Acknowledgments

Appendixes

A. GRADIENT AND HESSIAN

Recall that the loss function is $\ell(\boldsymbol{\theta}) = \sum_i \ell_i(\boldsymbol{\theta})$,

$$\ell_i(\boldsymbol{\theta}) = w_i \left\{ \left(\frac{\zeta}{z} - 1 \right) \log \frac{\zeta}{z} + \left(\frac{1-\zeta}{1-z} - 1 \right) \log \frac{1-\zeta}{1-z} \right\}, \qquad \zeta_i = c_i + D_{i\cdot}\,\boldsymbol{\theta}.$$

First note that $\frac{\partial \zeta_i}{\partial \theta_h} = \frac{\partial}{\partial \theta_h} \sum_j D_{ij}\theta_j = D_{ih}$. Next we prepare derivatives concerning $r_{0i} := \frac{\zeta_i}{z_i}$ and $r_{1i} := \frac{1-\zeta_i}{1-z_i}$:

$$\frac{\partial r_{0i}}{\partial \theta_h} = \frac{1}{z_i} \frac{\partial \zeta_i}{\partial \theta_h} = \frac{D_{ih}}{z_i} \qquad\qquad \frac{\partial r_{1i}}{\partial \theta_h} = -\frac{1}{1-z_i} \frac{\partial \zeta_i}{\partial \theta_h} = -\frac{D_{ih}}{1-z_i}$$

$$\frac{\partial}{\partial \theta_h} \log r_{0i} = \frac{1}{r_{0i}} \frac{\partial r_{0i}}{\partial \theta_h} = \frac{D_{ih}}{z_i r_{0i}} \qquad\qquad \frac{\partial}{\partial \theta_h} \log r_{1i} = \frac{1}{r_{1i}} \frac{\partial r_{1i}}{\partial \theta_h} = -\frac{D_{ih}}{(1-z_i) r_{1i}}$$

$$\frac{\partial}{\partial \theta_h} \frac{1}{r_{0i}} = -\frac{1}{r_{0i}^2} \frac{\partial r_{0i}}{\partial \theta_h} = -\frac{D_{ih}}{z_i r_{0i}^2} \qquad\qquad \frac{\partial}{\partial \theta_h} \frac{1}{r_{1i}} = -\frac{1}{r_{1i}^2} \frac{\partial r_{1i}}{\partial \theta_h} = \frac{D_{ih}}{(1-z_i) r_{1i}^2}$$

Then $\ell_i(\boldsymbol{\theta})$ may be differentiated componentwise:

$$
\begin{aligned}
\frac{\partial}{\partial \theta_h}(1 - r_{0i}) \log r_{0i} &= \frac{\partial r_{0i}}{\partial \theta_h} \log r_{0i} + (r_{0i} - 1) \frac{\partial \log r_{0i}}{\partial \theta_h} \\
&= \frac{D_{ih}}{z_i} \log r_{0i} + (r_{0i} - 1) \frac{D_{ih}}{z_i r_{0i}} \\
&= \frac{D_{ih}}{z_i} \left(1 - \frac{1}{r_{0i}} + \log r_{0i} \right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial \theta_h}(1 - r_{1i}) \log r_{1i} &= \frac{\partial r_{1i}}{\partial \theta_h} \log r_{1i} + (r_{1i} - 1) \frac{\partial \log r_{1i}}{\partial \theta_h} \\
&= -\frac{D_{ih}}{1 - z_i} \log r_{1i} + (r_{1i} - 1) \left(-\frac{D_{ih}}{(1 - z_i) r_{1i}} \right) \\
&= -\frac{D_{ih}}{1 - z_i} \left(1 - \frac{1}{r_{1i}} + \log r_{1i} \right)
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial}{\partial \theta_h} \{ (1 - r_{0i}) \log r_{0i} + (1 - r_{1i}) \log r_{1i} \} \\
=& \frac{D_{ih}}{z_i} \left(1 - \frac{1}{r_{0i}} + \log r_{0i} \right) - \frac{D_{ih}}{1 - z_i} \left(1 - \frac{1}{r_{1i}} + \log r_{1i} \right) \\
=& D_{ih} \left\{ \frac{1}{z_i} \left(1 - \frac{1}{r_{0i}} + \log r_{0i} \right) - \frac{1}{1 - z_i} \left(1 - \frac{1}{r_{1i}} + \log r_{1i} \right) \right\}
\end{aligned}
$$

The first derivative of the individual loss function is:

$$
\begin{aligned}
\frac{\partial}{\partial \theta_h} \ell_i(\boldsymbol{\theta}) &= \frac{\partial}{\partial \theta_h} w_i \{ (1 - r_{0i}) \log r_{0i} + (1 - r_{1i}) \log r_{1i} \} \\
&= w_i D_{ih} \left\{ \frac{1}{z_i} \left(1 - \frac{1}{r_{0i}} + \log r_{0i} \right) - \frac{1}{1 - z_i} \left(1 - \frac{1}{r_{1i}} + \log r_{1i} \right) \right\}.
\end{aligned}
$$

The second derivative is:

$$
\frac{\partial^2}{\partial\theta_h \partial\theta_k}\ell_i(\boldsymbol{\theta})
$$

$$
=\frac{\partial}{\partial\theta_k} w_i D_{ih}\left\{\frac{1}{z_i}\left(1-\frac{1}{r_{0i}}+\log r_{0i}\right)-\frac{1}{1-z_i}\left(1-\frac{1}{r_{1i}}+\log r_{1i}\right)\right\}
$$

$$
=w_i D_{ih}\left\{\frac{1}{z_i}\left(-\frac{\partial}{\partial\theta_k}\frac{1}{r_{0i}}+\frac{\partial}{\partial\theta_k}\log r_{0i}\right)-\frac{1}{1-z_i}\left(-\frac{\partial}{\partial\theta_k}\frac{1}{r_{1i}}+\frac{\partial}{\partial\theta_k}\log r_{1i}\right)\right\}
$$

$$
=w_i D_{ih}\left\{\frac{1}{z_i}\left(\frac{D_{ik}}{z_i r_{0i}^2}+\frac{D_{ik}}{z_i r_{0i}}\right)-\frac{1}{1-z_i}\left(-\frac{D_{ik}}{(1-z_i)r_{1i}^2}-\frac{D_{ik}}{(1-z_i)r_{1i}}\right)\right\}
$$

$$
=w_i D_{ih} D_{ik}\left\{\frac{1}{z_i^2 r_{0i}}\left(1+\frac{1}{r_{0i}}\right)+\frac{1}{(1-z_i)^2 r_{1i}}\left(1+\frac{1}{r_{1i}}\right)\right\}
$$

$$
=w_i D_{ih} D_{ik}\left(\frac{1+r_{0i}}{z_i^2 r_{0i}^2}+\frac{1+r_{1i}}{(1-z_i)^2 r_{1i}^2}\right)
$$

$$
=w_i D_{ih} D_{ik}\left(\frac{1+r_{0i}}{\zeta_i^2}+\frac{1+r_{1i}}{(1-\zeta_i)^2}\right).
$$

Summing up the terms and writing in matrix form, the gradient is:

$$
\mathbf{g}=\left[\begin{array}{c}\vdots\\\sum_i w_i D_{ih}\left\{\frac{1}{z_i}\left(1-\frac{1}{r_{0i}}+\log r_{0i}\right)-\frac{1}{1-z_i}\left(1-\frac{1}{r_{1i}}+\log r_{1i}\right)\right\}\\\vdots\end{array}\right]
$$

$$
=D'W\left[\begin{array}{c}\vdots\\\frac{1}{z_i}\left(\frac{\zeta_i-z_i}{\zeta_i}+\log\frac{\zeta_i}{z_i}\right)-\frac{1}{1-z_i}\left(\frac{z_i-\zeta_i}{1-\zeta_i}+\log\frac{1-\zeta_i}{1-z_i}\right)\\\vdots\end{array}\right]
$$

and the Hessian is

$$
H=\left[\begin{array}{ccc}\ddots & & \\ & \sum_i w_i D_{ih} D_{ik}\left(\frac{1+r_{0i}}{\zeta_i^2}+\frac{1+r_{1i}}{(1-\zeta_i)^2}\right) & \\ & & \ddots\end{array}\right]=
$$

$$
=D'W\left[\begin{array}{ccc}\ddots & & \\ & \frac{z_i+\zeta_i}{z_i\zeta_i^2}+\frac{(1-z_i)+(1-\zeta_i)}{(1-z_i)(1-\zeta_i)^2} & \\ & & \ddots\end{array}\right]D.
$$

B. GEOMETRICAL MEANING

A geometrical meaning of the semilog loss (4) is illustrated in Figure 7: the rectangular area delimited by the two dotted lines ($r = 1$ vertical and $\log r = 0$ horizontal) and the two solid lines ($r = 2$ vertical and $\log r = \log 2$ horizontal) equals $(r - 1) \log r$ for $r = 2$. So Figure 1 is the graph for such rectangular area with respect to r, as illustrated in Figure 7.
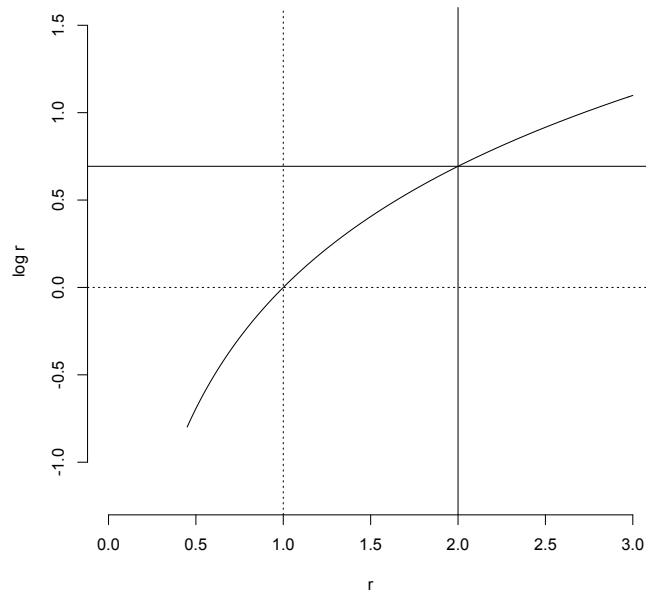

Fig. 7. Geometrical meaning of the semilog loss.
The rectangular area $(r - 1) \log r$ when $r = 2$

Figure 8 illustrates the meaning of the loss function (5); its interpretation is basically the same as with Figure 7. The vertical line in dotdash shows $z := \frac{1}{3}$ where the solid curve $\log \frac{\zeta}{z}$ and the dashed curve $\log \frac{1-\zeta}{1-z}$ both hit zero, the horizontal dotdash line. Now let $\zeta := \frac{7}{10}$ which is the vertical dotted line. Its corresponding values of $\log \frac{\zeta}{z}$ and $\log \frac{1-\zeta}{1-z}$ are shown by the upper and the lower horizontal dotted lines, respectively. The value of the first term in the right hand side of (5) is given by the rectangular area delimited by the upper horizontal dotted line, the vertical dotted line, and the horizontal and the vertical dotdash lines. Similarly, the value of the second term is given by the rectangular area delimited by the lower horizontal dotted line, the vertical dotted line, and the horizontal and the vertical dotdash lines. Hence the value of (5) is the rectangular area delimited by the three dotted lines and the

vertical dotdash line. When ζ moves closer to $z = \frac{1}{3}$ shown by the vertical dotdash line, the rectangular area shrinks to zero. If ζ moves further to the left beyond z the area again takes a positive value.
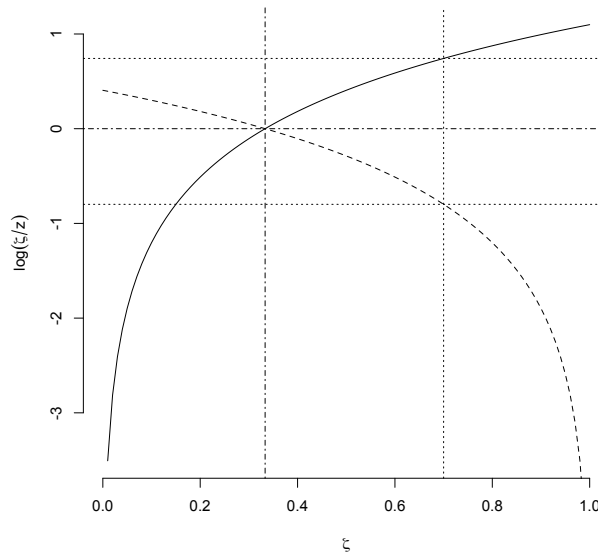

Fig. 8. Geometrical meaning of the proposed loss function.
$z = \frac{1}{3}$ (dotdash), $\zeta = \frac{7}{10}$ (dotted), $\log \frac{\zeta}{z}$ (solid), $\log \frac{1-\zeta}{1-z}$ (dashed)

REFERENCES

Campbell R.C., Hill R.C., 2006: Imposing Parameter Inequality Restrictions Using the Principle of Maximum Entropy. *Journal of Statistical Computation and Simulation* 76, 985–1000, see also http://www.bus.lsu.edu/academics/economics/faculty/chill/personal/maxent.htm.

Golan A., Judge G., Miller D., 1996: *Maximum Entropy Econometrics*. Wiley.

Herrnstein R.J., Rachlin H., Laibson D.I., 1997: *The Matching Law: Papers in Psychology and Economics*. Harvard University Press.

Layard R., 2006: *Happiness : Lessons from a New Science*. Penguin Books.

McCullagh P., Nelder J. A., 1989: *Generalized Linear Models*, Second Edition. Chapman & Hall.

Prelec D., January 1984: The Assumptions Underlying the Generalized Matching Law. *Journal of the Experimental Analysis of Behavior* 41 (1), 101–107.

Wichers R., 1996: *A Theory of Individual Behavior.* Academic Press.

Yoneda K., December 2006: A Parallel to the Least Squares for Positive Inverse Problems. *Journal of the Operations Research Society of Japan* 49 (4), 279–289.