Having Fun With A New Loss Function

Jan de Leeuw

First created on June 26, 2022. Last update on June 27, 2022

Abstract

The abstract

1 Introduction

In this paper a *loss function* is a real-valued non-negative function σ on $\mathbb{X}\otimes\mathbb{X}$, with $\mathbb{X}$ some subset of $\mathbb{R}^n$, and with $\sigma(x,y)=0$ if and only if $x=y$.

A loss function is *additive* if there is a real-valued non-negative *base function* f and positive weights w_i such that $\sigma(x,y)=\sum_{i=1}^n w_i f(x_i,y_i)$. This implies, of course, that $\sigma_i(x_i,y_i)=0$ if and only if $x_i=y_i$.

Loss functions are used to estimate *parameters*. This means that we assume x is the *model*. The model $x=\xi(\theta)$ is a function of $\theta\in\Theta$, where Θ is some subset of $\mathbb{R}^p$ and $\xi(\Theta)$ is a subset of $\mathbb{X}$. We want to find θ such that $F(\theta)$ approximates the *data* y. In most applications we are interested in the data are fixed, and it makes sense to define $\sigma(\theta):=\phi(\xi(\theta),y)$.

Thus we are interested in computing

$$\sigma_\star = \inf_{\theta\in\Theta}\sigma(\theta). \tag{1}$$

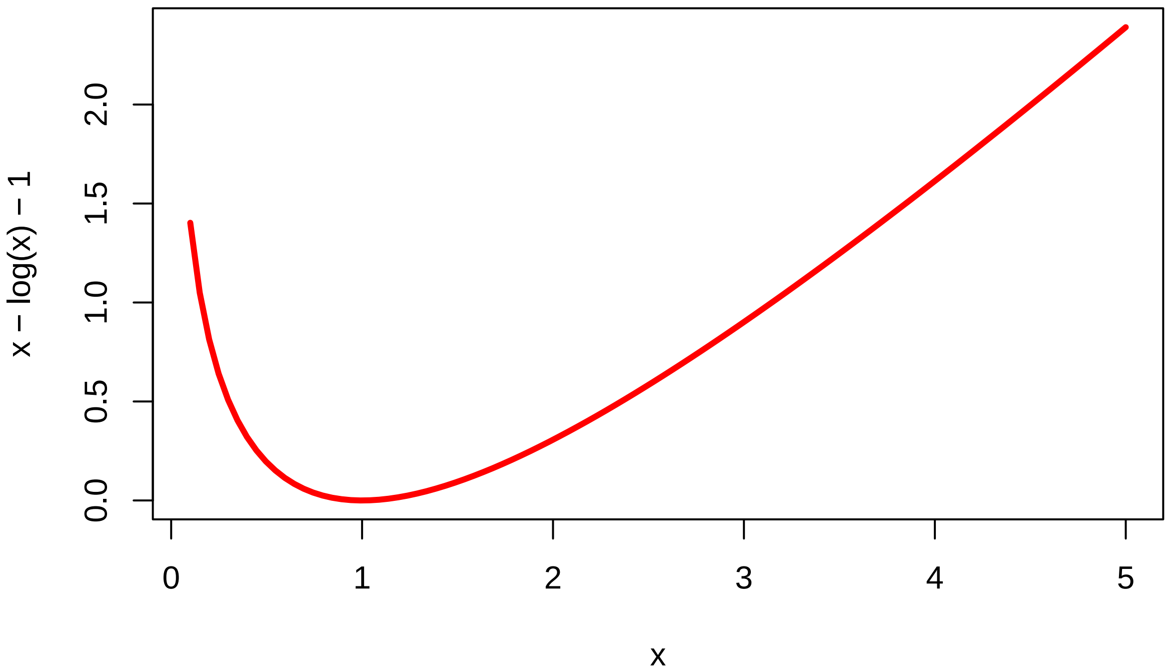
Since $\sigma\geq 0$

$$\hat\theta := \operatorname*{argmin}_{\theta\in\Theta}\sigma(\theta)=\{\theta\mid\sigma_\star=\sigma(\theta)\}. \tag{2}$$

Now if I say that we will discuss a *new* loss function, I obviously mean "new for me". I will mention some earlier work in which similar or even identical loss functions have appeared, but I am sure my literature review is far from complete.

2 Base Function

To construct the loss function we will discuss and apply in this paper we start with the function g, defined on the positive real axis by $g(x)=x-\log(x)-1$. $\sigma(x)\geq 0$ on the open interval $(0,+\infty)$, with $\sigma(x)=0$ if and only if $x=1$. Since log is strictly concave σ is strictly convex. At $x=0$ there is a vertical asymptote, i.e. $\lim_{x\downarrow 0}\sigma(x)=+\infty$.



The partial derivative is $\mathcal{D}\sigma(x) = 1 - \frac{1}{x}$ and thus $\mathcal{D}\sigma$ strictly increases with a horizontal asymptote $\lim_{x\uparrow\infty}\sigma'(x) = 1$ and with a vertical asymptote $\lim_{x\downarrow 0}\sigma'(x) = -\infty$.

The second partial derivative is $\mathcal{D}^2\sigma(x) = \frac{1}{x^2}$, which is postive and strictly decreases to its horizontal asymptote of zero. More generally, for $s \geq 2$ we have $\mathcal{D}^s\sigma(x) =) = (-1)^s(s-1)!\frac{1}{x^s}$. Thus $\mathcal{D}^s\sigma$ is strictly increasing and stricty concave for s odd and strictly decreasing and strictly convex for s even.

2.0.1 Base Loss Function

$$\sigma(x,y) = \frac{x}{y} - \log\frac{x}{y} - 1$$

ML connection

Swain (1975)

McDonald (1979)

$$\{\log x + \frac{y}{x}\} - \min_x\{\log x + \frac{y}{x}\} = (\frac{y}{x} - \log\frac{y}{x} - 1)$$

$$\sigma(x(\theta),y) = \frac{x(\theta)}{y} - \log\frac{x(\theta)}{y} - 1$$

$$\mathcal{D}_1\sigma(x(\theta),y) = \{\frac{1}{y} - \frac{1}{x(\theta)}\}\mathcal{D}x(\theta)$$

Note that if $x(\theta) < 0$ the partials are defined, even if the function is not.

$$\mathcal{D}_{11}\sigma(x(\theta),y)=\{\frac{1}{y}-\frac{1}{x(\theta)}\}\mathcal{D}^2x(\theta)+\frac{1}{x^2(\theta)}\mathcal{D}x(\theta)\mathcal{D}x(\theta)'$$

Thus if x is convex then σ is convex in θ on $\{\theta \mid x(\theta)\geq y\}$. Also if f is linear then σ is convex on $\{\theta \mid f(\theta)>0\}$.

$$\sigma(y,x(\theta))=\frac{y}{x(\theta)}-\log\frac{y}{x(\theta)}-1$$

$$D=-\frac{y}{x^2(\theta)}+\frac{1}{x(\theta)}=\frac{1}{x(\theta)}\left\{\frac{x(\theta)-y}{x(\theta)}\right\}$$

$$D^2=2\frac{y}{x^3(\theta)}-\frac{1}{x^2(\theta)}$$

$$f(\theta)=\sum_{i=1}^n\{f_i(\theta)-\log f_i(\theta)-1\}$$

3 Applications

3.1 Multiplicative Regression

$$f(\alpha)=\sum_{i=1}^n w_i\left\{\frac{\alpha x_i}{y_i}-\log\frac{\alpha x_i}{y_i}-1\right\}$$

$$\mathcal{D}f(\alpha)=\sum_{i=1}^n w_i\left\{\frac{x_i}{y_i}-\frac{1}{\alpha}\right\}$$

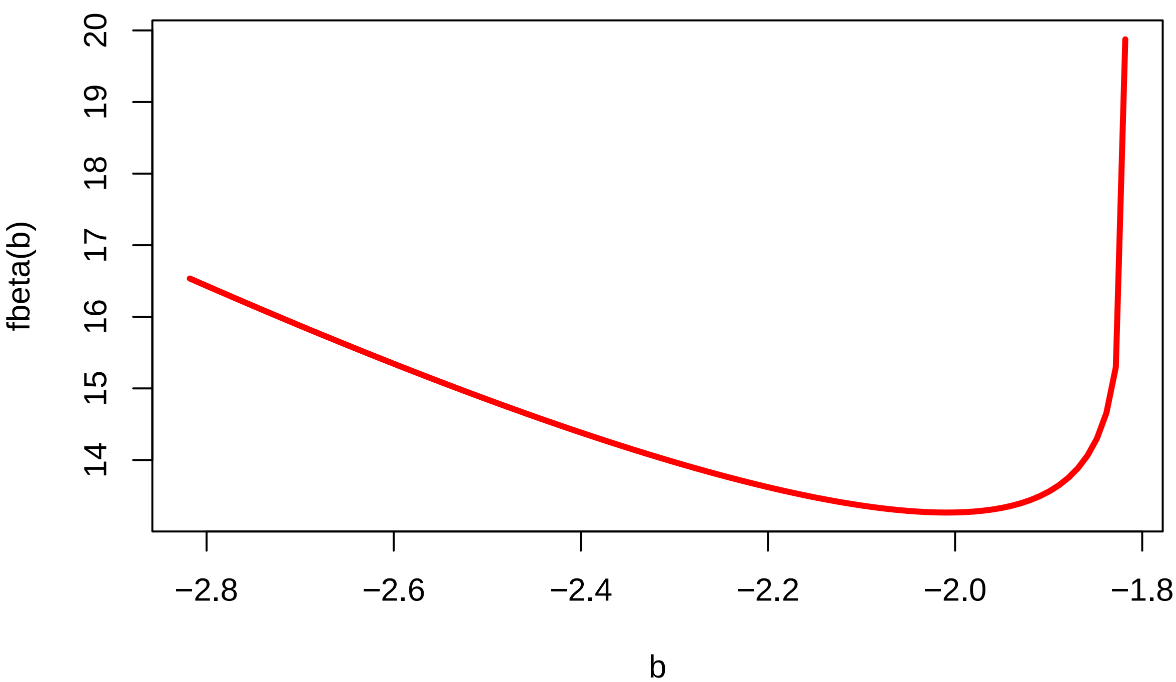
$$\hat{\alpha}=\mathcal{H}(\frac{y}{x},w)=\frac{\sum_{i=1}^n w_i}{\sum_{i=1}^n w_i\frac{x_i}{y_i}}.$$

3.1.1 Additive Regression

$$f(\beta)=\sum_{i=1}^n w_i\left\{\frac{x_i-\beta}{y_i}-\log\frac{x_i-\beta}{y_i}-1\right\}$$

The function f is defined only if $x_i-\beta>0$ for all i and thus for $\beta<x_{\min}$. In that region f is convex and we can apply Newton's method. But, as always with Newton, we have to be careful. We start, obviously, with $\beta<x_{\min}$, but the crux is to stay in the region. As f is very steep near the vertical asymptote, the minimum is likely to be close to $x_{\min}$. See figure ... If we start with choosing β too small then Newton will possibly take us to a $\beta>x_{\min}$, and we are in trouble. The algorithm resolves this by a combination of Newton and bisection. It makes sure that $\mathcal{D}f(\beta)$ is always positive, and thus it remains between the asymptote and the minimum and converges from the right.

```r
set.seed(12345)
x <- rnorm(10)
y <- rnorm(10) ^ 2
m <- min(x)
fbeta <- function (b) {
  v <- outer(x, b, "-")
  return(colSums(v)-colSums(log(v)) - 1)
}
b <- seq(m - 1, m - .0001, length = 100)
plot(b, fbeta(b), type = "l", col = "RED", lwd = 3)
```



$$\mathcal{D}f(\beta) = \sum_{i=1}^{n} \left\{ -\frac{1}{y_i} + \frac{1}{x_i - \beta} \right\}$$

$$\mathcal{D}^2 f(\beta) = \sum_{i=1}^{n} \frac{1}{(x_i - \beta)^2}$$

```r
source("newtonb.R")
newtonb(x, y, -3)
```

```
## itel     1 fold  207.294275 fnew  207.293960 bnew  -1.827421 gnew  0.065441 hnew  111
## itel     2 fold  207.293960 fnew  207.293960 bnew  -1.827426 gnew  0.000040 hnew  111
## itel     3 fold  207.293960 fnew  207.293960 bnew  -1.827426 gnew  0.000000 hnew  111
```

```
## $b
```

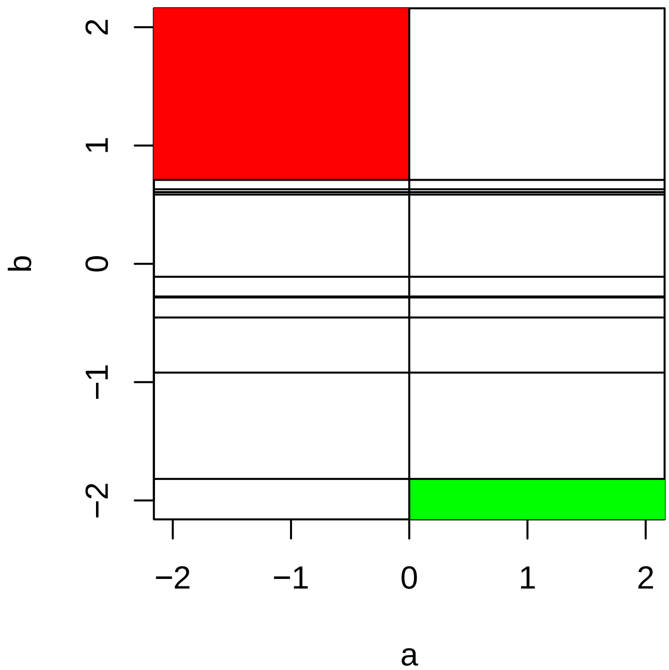
```
## [1] -1.827426376
##
## $f
## [1] 207.2939602
##
## $g
## [1] 1.522465487e-11
##
## $h
## [1] 11153.26338
##
## $itel
## [1] 3
```

3.1.2 Linear Regression

$$f(a,b)=\sum_{i=1}^n\left\{\frac{a(x_i-b)}{y_i}-\log\frac{a(x_i-b)}{y_i}-1\right\}$$

$a(x_i-b)\geq 0$ for all i iff either $a>0$ and $b<x_{\min}$ or $a<0$ and $b>x_{\max}$.

```
par(pty="s")
a <-seq(-2,+2,length=10)
plot(cbind(a,x), ylim = c(-2,2), ylab = "b", type = "n")
rect(-3, max(x), 0, 3, col = "RED", lwd = 0, density = -1)
rect(0, -3, 3, min(x), col = "GREEN", lwd = 0, density = -1)
for (i in 1:10) {
  abline(h = x[i])
}
abline(v = 0)
```



$$\mathcal{D}_1 f(a,b) = \sum_{i=1}^n \left\{\frac{x_i - b}{y_i}\right\} - \frac{n}{a}, \tag{3}$$

$$\mathcal{D}_2 f(a,b) = \sum_{i=1}^n \left\{\frac{-a}{y_i} + \frac{1}{x_i - b}\right\}. \tag{4}$$

$$\mathcal{D}_{11} f(a,b) = \frac{n}{a^2}, \tag{5}$$

$$\mathcal{D}_{22} f(a,b) = \sum_{i=1}^n \frac{1}{(x_i - b)^2}, \tag{6}$$

$$\mathcal{D}_{21} f(a,b) = \mathcal{D}_{12} f(a,b) = \sum_{i=1}^n \frac{1}{y_i}. \tag{7}$$

Extend to linear and polynomial

3.1.3 Monotone Regression

In Monotone Regression (for a weak linear order) we minimize

$$\sum_{i=1}^n w_i\{\frac{x_i}{y_i} - \log\frac{x_i}{y_i} - 1\} \tag{8}$$

over all $x_1 \leq x_2 \leq \cdots \leq x_{n-1} \leq x_n$.

Lemma 3.1 (Monotone Regression Lemma)**.** *Suppose $n = 2$ and we require $x_1 \leq x_2$. If $y_1 \leq y_2$ then $\hat{x}_1 = y_1$ and $\hat{x}_2 = y_2$. If $y_i \geq y_2$ then $\hat{x}_1 = \hat{x}_2 = \mathcal{H}(y_1, y_2; w_1, w_2)$*

Proof. We must minimize the convex function σ on the cone $0 < x_1 \leq x_2$. There are only two possibilities: either (y_1, y_2) is inside the cone, or it is outside. If it is inside the cone the gradient must vanish, which means $(\hat{x}_1, \hat{x}_2) = (y_1, y_2)$. If (y_1, y_2) is outside the cone, i.e. $0 < y_2 < y_1$, we project on the boundary line $x_1 = x_2$, which produces the weighted harmonic mean of y_1 and y_2. □

Theorem 3.1 (PAVA Theorem)**.** *If $y_i > y_{i+1}$ then $\hat{x}_i = \hat{x}_{i+1}$.*

Proof. If $\hat{x}_i < \hat{x}_{i+1}$ then the Monotone Regression Lemma shows that merging improves the fit. And of course the merged values are still feasible. □

From theorem 3.1 it follows we can compute the monotonic regression by a Pooled Adjacent Violaters Algorithm, or PAVA (J. De Leeuw, Hornik, and Mair (2009)). It also follows that the reasoning in J. De Leeuw (1977) applies for the case in which the y_i have ties.

The PAVA algorithm looks for a violation $y_i > y_{i+1}$. It then uses this violation, and theorem @3.1, to reduce the problem from one of size n to one of size $n-1$. w_i and w_{i+1} are replaced by $w_i + w_{i+1}$ and y_i and y_{i+1} by $\mathcal{H}(y_i, y_{i+1}; w_i, w_{i+1})$. We then continue in the same way with the smaller problem until we are ultimately left with a vector of merged elements that are in the correct order.

We give a simple example where we proceed strictly from left to right. All weights are one, which means that after merging the weights are the sizes of the blocks. The matrices below have the merged y_1 in the first row and the block sizes in the second row. If the weights are arbitrary positive numbers we maintain three rws: block values, block weights, and block sizes. Efficient ways of merging and selecting which violations to use are discussed by Busing (2022).

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## [1,]    2    1    3    1    1    5    3
## [2,]    1    1    1    1    1    1    1
```

```
##             [,1] [,2] [,3] [,4] [,5] [,6]
## [1,] 1.333333333    3    1    1    5    3
## [2,] 2.000000000    1    1    1    1    1
```

```
##             [,1] [,2] [,3] [,4] [,5]
## [1,] 1.333333333  1.5    1    5    3
## [2,] 2.000000000  2.0    1    1    1
```

```
##             [,1]        [,2] [,3] [,4]
## [1,] 1.333333333 1.285714286    5    3
## [2,] 2.000000000 3.000000000    1    1
```

```
##             [,1] [,2] [,3]
## [1,] 1.304347826    5    3
## [2,] 5.000000000    1    1
```

```
##             [,1] [,2]
## [1,] 1.304347826 3.75
## [2,] 5.000000000 2.00
```

Now the two remaining elements are in the correct order and we expand the solution to its original length, using the block sizes.

```
## [1] 1.304347826 1.304347826 1.304347826 1.304347826 1.304347826 3.750000000
## [7] 3.750000000
```

Just to make sure we did no go astray, we check the Karush-Kuhn-Tucker conditions for minimizing $\sigma(x, y)$ over x satisfying $Ax \leq 0$, where A is

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## [1,]    1   -1    0    0    0    0    0
## [2,]    0    1   -1    0    0    0    0
## [3,]    0    0    1   -1    0    0    0
## [4,]    0    0    0    1   -1    0    0
## [5,]    0    0    0    0    1   -1    0
## [6,]    0    0    0    0    0    1   -1
```

The gradient $\nabla\sigma(\hat{x})$ is

```
## [1] -0.26666666667  0.23333333333 -0.43333333333  0.23333333333  0.23333333333
## [6] -0.06666666667  0.06666666667
```

and the Lagrange multipliers $\hat{\lambda}$ are the solution of $A'\lambda = \nabla\sigma(\hat{x})$, which gives

```
## [1] 0.26666666667 0.03333333333 0.46666666667 0.23333333333 0.00000000000
## [6] 0.06666666667
```

The right-hand sides $\hat{r} = A\hat{x}$ are

```
## [1]  0.000000000  0.000000000  0.000000000  0.000000000 -2.445652174
## [6]  0.000000000
```

The KKT conditions $\hat{\lambda} \geq 0$ and $\hat{r} \leq 0$ are satisfied, and so is strict complementarity which requires that for each i either $\hat{r}_i = 0$ or $\hat{\lambda}_i = 0$ (but not both). Thus we have indeed computed the unique minimum of σ over the cone $x_1 \leq \cdots \leq x_n$.

3.2 Factor Analysis

Let me first report a happy coincidence. In a really excellent, and unfortunately somewhat neglected paper, Swain (1975) discussed several loss functions for (random orthogonal exploratory) factor analysis with the same statistical properties as maximum likelihood. Minimizing those loss function gives estimates that have the same asymptotic standard errors and chi-square tests as obtained from the maximum likelihood estimates. All these loss functions are functions of the eigenvalues θ_i of $S^{-1}\Sigma$, where S is the sample covarinace matrix and $\Sigma = AA' + \Delta^2$ is the factor analysis mode. In particular, Swain shows that the maximum likelihood estimation corresoinds with minimizing

$$\sum_{i=1}^n \left\{\frac{1}{\theta_i} + \log(\theta_i) - 1\right\},$$

which is obviously the same as minimizing

$$\sigma(\lambda) = \sum_{i=1}^n \left\{\frac{\lambda_i}{1} - \log(\frac{\lambda_i}{1}) - 1\right\},$$

where the λ_i are the eigenvalues of $\Sigma^{-1}S$. Thus maximum likelihood estimation in factor analysis is a special case of minimizing our new fun loss function σ.

The original motivation for the new loss function, however, comes from De Leeuw J (n.d.), and is more closely related to Mc Donald's maximum likelihood ratio estimate in (fixed orthogonal exploratory) factor analysis (McDonald (1979)).

Suppose $\underline{z}_i$ are independent normallly distributed random variables, with

$$\underline{z}_i \sim \mathcal{N}(y_i - f_i(\theta),\ \delta_i^2(\eta)).$$

Both means and variances depend on parameters, respectively θ and η. The the deviance (two times the negative log-likelihood) is

$$\Delta(\theta, \eta) = \sum_{i=1}^n \left\{\log \delta_i^2(\eta) + \frac{(y_i - f_i(\theta))^2}{\delta_i^2(\eta)}\right\}$$

Presumably we want to minimize this over θ and η, but this runs into problems if there is at least one index k, a $\hat{\theta}$, and a $\hat{\eta}$ for which $y_k - f_k(\theta) = 0$ and $\lim_{\eta\rightarrow\hat{\eta}} \delta_k^2(\eta) = 0$. If that is the case then $\inf_\theta \inf_\eta \Delta(\theta, \eta) = -\infty$ and the minimum (and thus the maximum lilkeihood estimate) does not exist (Anderson and Rubin (1956)).

McDonald (1979) proposes to subtract the minimum deviance over the unconstrained δ_i for given residuals $y_i - f_(theta)$ from the deviance in equation … Now

$$\min_\delta \Delta(\theta, \delta) = \min_\delta \sum_{i=1}^n \left\{\log \delta_i^2 + \frac{(y_i - f_i(\theta))^2}{\delta_i^2}\right\} = \sum_{i=1}^n \left\{\log(y_i - f_i(\theta))^2 + 1\right\},$$

and thus

$$\Delta(\theta,\eta)-\min_\delta \Delta(\theta,\delta)=\sum_{i=1}^n\left\{\frac{(y_i-f_i(\theta))^2}{\delta_i^2(\eta)}-\log\frac{(y_i-f_i(\theta))^2}{\delta_i^2(\eta)}-1\right\}=\sigma((y_i-f_i(\theta))^2,\ \delta_i^2(\eta))$$

Subtracting the term $\min_\delta \Delta(\theta,\delta)$ works as a barrier pennaty function, preventing that $y_i-f_i(\theta)=0$. Now I am not sure about the statistical optimality properties of maximum likelihood ratio estimates, but in this case they lead us to our fun loss function and that's all I care about here. This approach is analyzed, with examples, in more detail in De Leeuw J (n.d.). Clearly it can be used to fit not just fixed score factor analysis, but a wide variety of mean/variance structure models.

3.3 MDS/Unfolding

Another motivation for looking at loss function σ is that its zero-avoidance properties may be helpful in multidimensional scaling, in particular in unfolding, and even more in particular in non-metric unfolding. We have developed the appropriate form of monotone regression in section ..., and consequently we can use alternating minimization to turn metric MDS into non-metric MDS.

Notation: $A_{ij}:=I_p\otimes(e_i-e_j)(e_i-e_j)'$ and $\text{vec}(X)$. Then $d_{ij}^2(X)=x'A_{ij}^x$ smacof

$$f(x)=\frac12\sum_{i=1}^n\sum_{j=1}^n w_{ij}\left\{\frac{x'A_{ij}x}{\delta_{ij}^2}-\log\frac{x'A_{ij}x}{\delta_{ij}^2}-1\right\}=$$
$$\frac12 x'Sx-\frac12\sum_{i=1}^n\sum_{j=1}^n w_{ij}\log x'A_{ij}x+\frac12\sum_{i=1}^n\sum_{j=1}^n w_{ij}\log\delta_{ij}^2-\frac12\sum_{i=1}^n\sum_{j=1}^n w_{ij}, \quad (9)$$

with

$$S:=\sum_{i=1}^n\sum_{j=1}^n w_{ij}\frac{1}{\delta_{ij}^2}A_{ij}. \quad (10)$$

The first derivatives are

$$\mathcal{D}f(x)=Sx-\sum_{i=1}^n\sum_{j=1}^n w_{ij}\frac{1}{d_{ij}^2(x)}A_{ij}x=(S-T(x))x,$$

with

$$T(x):=\sum_{i=1}^n\sum_{j=1}^n w_{ij}\frac{1}{d_{ij}^2(x)}A_{ij}. \quad (11)$$

For the second derivatives we find

$$\mathcal{D}^2f(x)=S-\sum_{i=1}^n\sum_{j=1}^n w_{ij}\frac{1}{d_{ij}^2(x)}\left\{A_{ij}-2\frac{1}{d_{ij}^2(x)}A_{ij}xx'A_{ij}\right\}=S-T(x)+2U(x)=S-(T(x)-U(x))+U(x),$$

with

$$U(x)=\sum_{i=1}^n\sum_{j=1}^n w_{ij}\frac{1}{d_{ij}^4(x)}A_{ij}. \tag{12}$$

Note that S is positive semi-definite and doubly-centered. if $w_{ij}\delta_{ij}$ is irreducible then S has rank $n-1$, with only the vectors propertional to e in its null-space. In the same way $T(x)$ and $U(x)$ are positive semi-definite and doubly-centered for all x, and so is $T(x)-U(x)$.

$$y'\mathcal{D}^2f(x)y=y'Sy-\sum\sum w_{ij}\frac{y'A_{ij}y}{d_{ij}^2(x)}\{1-2\frac{(x'A_{ij}y)^2}{x'A_{ij}x\times y'A_{ij}y}\}$$

Thus

$$S-T(x)\lesssim\mathcal{D}^2f(x)\lesssim S+T(x)$$

Note that both S and $T(x)$ are block-diagonal, but $U(x)$ is a full matrix with non-zero off-diagonal blocks. Also if $S-T(x)\gtrsim 0$, i.e. $S^+T(x)\leq I$ then $\mathcal{D}^2f(x)\gtrsim 0$.

try

$$x^+=x-(S+T(x))^{-1}(S-T(x))x=(S+T(x))^{-1}\{(S+T(x))-(S-T(x))\}=2(S+T(x))^{-1}T(x)$$

$$x^+=x-(S+U(x))^{-1}(S-T(x))x=(S+U(x))^{-1}\{(S+U(x))-(S-T(x))\}x==(S+T(x))^{-1}(U(x)+T(x))x$$

$$x^+=x-(S-T(x)+2U(x))^{-1}(S-T(x))x=(S-T(x)+2U(x))^{-1}\{(S-T(x)+2U(x))-(S-T(x))\}x=2(S-T$$

Check with numDeriv

3.4 Log-linear

$$\sum_{i=1}^n\frac{\pi_i(\theta)}{p_i}-\log\frac{\pi_i(\theta)}{p_i}-1$$

Independence

$$\sum_{i=1}^n\sum_{j=1}^m\left\{\frac{\alpha_i\beta_j}{p_{ij}}-\log\frac{\alpha_i\beta_j}{p_{ij}}-1\right\}$$

$$\alpha_i=\mathcal{H}(\frac{p_{i\bullet}}{\beta})$$

$$\mathcal{L}=\sum_i\sum_j\sum_k\left\{\frac{\alpha_{ij}\beta_{ik}\gamma_{jk}}{p_{ijk}}-\log\frac{\alpha_{ij}\beta_{ik}\gamma_{jk}}{p_{ijk}}-1\right\}$$

$$\frac{\partial\mathcal{L}}{\partial\alpha_{ij}}=\sum_k\frac{\beta_{ik}\gamma_{jk}}{p_{ijk}}-\frac{1}{\alpha_{ij}}$$

instead of $\mathcal{L}(\pi,p)$ also $\mathcal{L}(p,\pi)$

$$\mathcal{L}=\sum_i\sum_j\sum_k\{p_{ijk}\alpha_{ij}\beta_{ik}\gamma_{jk}-\log p_{ijk}\alpha_{ij}\beta_{ik}\gamma_{jk}-1\}$$

$$\sum_j\sum_k p_{ijk}\beta_{ik}\gamma_{jk}-\frac{1}{\alpha_{ij}}=0$$

References

Anderson, T. W., and H. Rubin. 1956. "Statistical Inference in Factor Analysis." In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, edited by J. Neyman, V:111–50. Berkeley; Los Angeles: University of California Press.

Busing, F. M. T. A. 2022. "Monotone Regression: A Simple and Fast O(n) PAVA Implementation." *Journal of Statistical Software* 102 (Code Snippet 1).

De Leeuw, J. n.d. "Factor Analysis, Correspondence Analysis, ANOVA." https://jansweb.netlify.app/publication/deleeuw-e-22-c.

De Leeuw, J. 1977. "Correctness of Kruskal's Algorithms for Monotone Regression with Ties." *Psychometrika* 42: 141–44.

De Leeuw, J., K. Hornik, and P. Mair. 2009. "Isotone Optimization in R: Pool-Adjacent-Violators Algorithm (PAVA) and Active Set Methods." *Journal of Statistical Software* 32 (5): 1–24.

McDonald, R. P. 1979. "The Simultaneous Estimation of Factor Loadings and Scores." *British Journal of Mathematical and Statistical Psychology* 32 (212–228).

Swain, A. J. 1975. "A Class of Factor Analysis Estimation Procedures with Common Asymptotic Sampling Properties." *Psychometrika* 40 (3): 315–36.