Inverse Problems in Statistics
Chapter 5: Computing the Estimates

Diego Colombo, Luca Gugelmann

21. May 2007

All statements in this short handout are provided for reference and without proof. For more details see [2].

1 Some convex optimization

Definition 1.1 (Convex cone) *A* finitely generated convex cone *in* $\mathbb{R}^k$ *is a subset* S *of* $\mathbb{R}^k$ *such that for given linearly independent vectors* $\nu^{(1)}, \nu^{(2)}, \ldots, \nu^{(m)}$ *in* $\mathbb{R}^k$ *(the* generators *of* S*) we have:*

$$s \in S \Leftrightarrow \exists\, \lambda_1, \lambda_2, \ldots \lambda_m \geq 0 \text{ such that: } s = \sum_{i=1}^{m} \lambda_i \nu^{(i)}$$

A special case of a finitely generated convex cone is the following:

$$\mathcal{C} = \left\{ s = (s1, \ldots, s_n) \in \mathbb{R}^n \mid s_1 \leq s_2 \leq \ldots \leq s_n \right\}$$

Lemma 1.2 *Suppose that* S *is a finitely generated convex cone in* $\mathbb{R}^k$ *and that* ϕ *is convex and continuously differentiable on* S*. Then* $\hat{s} \in argmin_{s \in S} \phi(s)$ *if and only if*

$$\hat{s} \in S, \forall i \in \{1, 2, \ldots, m\}\ \nabla\phi(\hat{s})^T \nu^{(i)} \geq 0, \text{and } \nabla\phi(\hat{s})^T \hat{s} = 0$$

2 Isotonic regression

Definition 2.1 (Isotonic function) *Given a set* M *with a partial order* $\preceq$*, a real-valued function* μ *on* M *is called isotonic if:*

$$x \preceq y \Rightarrow \mu(x) \preceq \mu(y)$$

In the special case $M = \mathbb{R}$*,* $\preceq$ *the usual order on the real numbers, an isotonic function is just a monotone function.*

The objective of regression analysis is to describe the behavior of a variate given one or more covariates. Formally, one is interested in an estimate of the function $\mu_0(x) = \mathrm{E}[\,Y \mid X = x\,]$ for some random vector $(X, Y) \in \mathbb{R} \times \mathbb{R}^p$. The objective of *isotonic regression* is the same, but we limit the function μ_0 to a given set of isotonic functions. Given a function μ which is a solution to a regression problem, one way of computing its equivalent in an isotonic setting is to just take the L^2-projection of μ into the given set of isotonic functions. Specifically we will look at the projection of estimates (in specific examples MLEs) into the space $\mathcal{C}$.

3 Weighted least squares projection onto $\mathcal{C}$

Mastering the following problem will prove extremely useful in the course of our discussion of the iterative convex minorant algorithm. Given a point $u \in \mathbb{R}^k$, how can one compute the (weighted) least squares projection of u into $\mathcal{C}$? More specifically, for:

$$\phi(x) = \sum_{i=1}^{k} (x_i - u_i)^2 w_i \quad \text{with } w_i \text{ the weights.}$$

we want to compute $\operatorname{argmin}_{y \in \mathcal{C}} \phi(y)$.
Applying Lemma 1.2, one can see that this problem has a nice geometrical solution. Consider the points,

$$P_0 = (0,0) \quad \text{and} \quad P_i = \Big(\sum_{j=1}^{i} w_j, \sum_{j=1}^{i} w_j u_j \Big),\ j \in \{1, \dots, k\}.$$

the least squares projection of u into $\mathcal{C}$ is now given by the left derivatives of the convex minorant of those points.

4 The iterative convex minorant algorithm

The optimization problem we want to solve is the following: given a function $\phi : \mathbb{R}^n \to (-\infty, \infty]$, compute:

$$\hat{\beta} = \operatorname*{argmin}_{\beta \in \mathcal{C}} \phi(\beta)$$

where ϕ satisfies the conditions:

Conditions 4.1 *i) ϕ is convex, continuous and attains its minimum over $\mathcal{C}$ at a unique point $\hat{x}$.*

ii) ϕ is continously differentiable on the set $\{x \in \mathbb{R}^n \mid \phi(x) < \infty\}$.

Since ϕ is an arbitrary function, we try to reduce this optimization problem to an easier case by taking a second order Taylor approximation of it, but instead of having D be the Hessian matrix, we allow it to be any positive definite diagonal matrix. Given a point $\gamma \in \mathcal{C}$ with $\phi(\gamma) < \infty$ we have:

$$\begin{aligned}
\phi(\beta) - \phi(\gamma) &= (\beta - \gamma)^T \nabla\phi(\gamma) + \frac{1}{2}(\beta - \gamma)^T D(\beta - \gamma) + o(\|\beta - \gamma\|) \\
&= c_\gamma + \frac{1}{2}(\beta - \gamma + D^{-1}\nabla\phi(\gamma))^T D(\beta - \gamma + D^{-1}\nabla\phi(\gamma)) + o(\|\beta - \gamma\|)
\end{aligned}$$

where c_γ does not depend on β. A natural choice at this point is to base the algorithm on this algorithmic map:

$$B(\gamma) = \underset{\beta \in \mathcal{C}}{\operatorname{argmin}} \frac{1}{2}(\beta - \gamma + D^{-1}\nabla\phi(\gamma))^T D(\beta - \gamma + D^{-1}\nabla\phi(\gamma))$$

that is, choose a starting point $\beta^{(0)}$, compute $\beta^{(1)} = B(\beta^{(0)})$ and so on, hoping for convergence. Note that since D is a diagonal matrix, the above optimization problem is equivalent to finding the weighted least squares projection of β in $\mathcal{C}$ where the weights are the entries on the diagonal of D, and therefore we can find a solution via the convex minorant as discussed before. This is the main idea behind the iterative convex minorant algorithm. Unfortunately there are situations where this does not converge. The approach is still useful however, as can be seen with the following lemma:

Lemma 4.2 *Let ϕ satisfy the conditions 4.1 and $\beta \in \mathcal{C} \setminus \{\hat{\beta}\}$ satisfy $\phi(\beta) < \infty$, then:*

$$\phi(\beta + \lambda(B(\beta) - \beta)) < \phi(\beta)$$

for all $\lambda > 0$ sufficiently small.

This lemma states that for a given β, $B(\beta)$ defines a *descent direction* along which one can find a β' such that $\phi(\beta') < \phi(\beta)$, and thus a way to find the minimizer of the convex function ϕ. Specifically, one can perform a binary search along the line from β to $B(\beta)$ to find such a β'. The iterative convex minorant algorithm with this addition is known as the modified iterative convex minorant algorithm, see Algorithm 1 at the end of this document for a pseudo-code description of it. Convergence of this is guaranteed by the following theorem:

Theorem 4.3 *Let the function $\phi : \mathbb{R}^n \to (-\infty, \infty]$ satisfy the conditions 4.1 and $\beta^{(0)} \in \mathcal{C}$ satisfy $\phi(\beta^{(0)}) < \infty$. Let the mapping $\beta \mapsto D(\beta)$ take values in the set of positive definite $n \times n$ diagonal matrices such that $\beta \to D(\beta)$ is continuous on the set:*

$$K = \left\{\beta \in \mathcal{C} \mid \phi(\beta) \leq \phi(\beta^{(0)})\right\}$$

Then modified iterative convex minorant algorithm converges to $\hat{\beta}$.

5 Application: double censoring

We will retrace the simulation performed in [1]. The setting for the double censoring problem has been presented in other talks before, so we just concentrate on the simulation. We generate samples of various sizes from $F(x) = \sqrt{x} I_{[0,1)}(x) + I_{[1,\infty)}$, and we choose (T, U) (the interval bounds) uniformly distributed on $\{(t, u) \in [0, 1]^2 \mid t < u\}$. From this data we compute $v_1 < v_2 < \ldots < v_l$ by taking $v_j = T_i$ if $X_i < T_i$; $v_j = T_i$, $v_{j+1} = U_i$ if $T_i < X_i < U_i$ and finally $v_j = U_i$ if $X_i > U_i$. Each index in $\{1, \ldots, l\}$ is added to one of the four sets I_1, I_{2a}, I_{2b} and I_3, depending on which one of the cases above holds for v_i. Also a mapping k from I_{2a} to I_{2b} is computed, linking each lower bound in I_{2a} to the corresponding upper bound in I_{2b}. The function to be minimized is:

$$\phi(\beta) = -\frac{1}{n}\Big(\sum_{i \in I_1} \log \beta_i + \sum_{i \in I_{2a}} \log(\beta_{k(i)} - \beta_i) + \sum_{i \in I_3} \log(1 - \beta_i)\Big)$$

The β_i are related to F by $\beta_i = F(v_i)$, therefore $\beta \in \mathcal{C}$ is ensured. Also $\beta_1 \geq 0$ and $\beta_l \leq 1$ need to be satisfied. This holds automatically if $1 \in I_1$ and $l \in I_3$, therefore we create 2 artificial points $v_1 = 0$ and $v_l = 1$ and add the corresponding indices to I_1 and I_3. Without these additional points the algorithm would not converge to a meaningful solution, and would choose $\beta_i > 1$ for some i, which could lead to a negative value in the logarithm.

References

[1] G. Jongbloed. The iterative convex minorant algorithm for nonparametric estimation. *J. Comput. Graph. Statist.*, 7:310–321, 1998.

[2] G. Jongbloed. Inverse problems in statistics, 1999.

Algorithm 1 Modified iterative convex minorant algorithm

Input:
$\eta > 0$: accuracy parameter
$\epsilon \in (0, 1/2)$: line search parameter
$\beta^{(0)} \in \mathcal{C}$: initial point satisfying $\phi(\beta^{(0)}) < \infty$

$\beta := \beta^{(0)}$
while $\left|\sum_{i=1}^n \beta_i \frac{\partial}{\partial \beta_i}\phi(\beta)\right|$ **or** $\left|\sum_{i=1}^n \frac{\partial}{\partial \beta_i}\phi(\beta)\right|$ **or** $\min_{1\leq j\leq n} \sum_{i=j}^n \frac{\partial}{\partial \beta_i} < -\eta$ **do**
  $\tilde{y} := \operatorname{argmin}_{y\in\mathcal{C}}(y - \beta + D(\beta)^{-1}\nabla\phi(\beta))^T D(\beta)(y - \beta + D(\beta)^{-1}\nabla\phi(\beta))$
  if $\phi(\tilde{y}) < \phi(\beta) + \epsilon\nabla\phi(\beta)^T(\tilde{y} - \beta)$ **then**
    $\beta := \tilde{y}$
  else
    $\lambda := 1; s := 1/2; z := \tilde{y}$
    while $\phi(z) < \phi(\beta) + (1-\epsilon)\nabla\phi(\beta)^T(z-\beta)$ (I) **or**
        $\phi(z) > \phi(\beta) + \epsilon\nabla\phi(\beta)^T(z-\beta)$ (II) **do**
      if (I) **then**
        $\lambda := \lambda + s$
      end if
      if (II) **then**
        $\lambda := \lambda - s$
      end if
      $z := \beta + \lambda(\tilde{y} - \beta)$
      $s := s/2$
    end while
    $\beta := z$
  end if
end while