Supra-Bayesian combination of probability distributions - part III

Vladimíra Sečkárová

14.05.2011

Situation

- we have: a random vector described by unknown probability mass function h
- aim is: to evaluate the estimate of h based on data D
- form of the estimate: $\hat{h} = \mathrm{E}_{\pi(h|D)}(h|D)$ (based on Kerridge inaccuracy)
- question 1: form of the posterior pdf $\pi(h|D)$
- question 2: form of the estimate based on used $\pi(h|D)$

Posterior pdf

- posterior pdf:
  - maximizes the entropy subject to dissimilarity constraints
  - constraint considered for j^{th} source is: $\mathrm{EK}(g_j, h) = \beta_j$
- aim is: to compute the Lagrangian of this optimization task
- Lagrangian: $L(.,.) = \mathrm{D}(\pi(h|D)||^{O}\pi(h|D)) - \log \frac{\Pi_i \Gamma(\nu_i)}{\Gamma(\nu_0)} - \sum_j \lambda_j \beta_j$
- ${}^{O}\pi(h|D)$ - pdf of Dirichlet distribution, $\lambda_j > 0$ - j^{th} Lagrange multiplier (representing the weights)
- estimate form: $\hat{h}(.) = \frac{1}{n+\sum_j \lambda_j} + \sum_j \frac{\lambda_j g_j(.)}{n+\sum_j \lambda_j}$
- question: choice of β_j
- question: evaluation of λs

Choice of βs

- the following holds: $\beta_j = \mathrm{EK}(g_j, h) \geq H(g_j)$
- each β_j is set as $H(g_j)$ (entropy of j^{th} source)
- possible reason: smaller value of βs - better performance in λs (in the sense of sources differentiability)
- with higher βs the values of Lagrange multipliers declines $\rightarrow$ with higher βs the weights come closer

Evaluation of λs

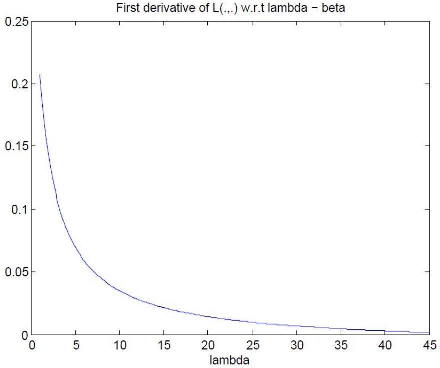
- we are looking for λ which minimizes the Lagrangian = we want to find λ_j such that first derivative w.r.t. each λ_j of Lagrangian equals zero
- formula for λ_j: $-\sum_i \psi(\nu_i) g_j + \psi(\nu_0) = \beta_j$
- where: $\nu_i = 1 + \sum_j \lambda_j g_j(x_i)$, $\nu_0 = n + \sum_j \lambda_j$
- we have a system of nonlinear equations:

$$-\begin{pmatrix} (\quad g_1 \quad) \\ \vdots \\ (\quad g_s \quad) \end{pmatrix} \begin{pmatrix} \psi(\nu_1) \\ \vdots \\ \psi(\nu_n) \end{pmatrix} + \begin{pmatrix} \psi(\nu_0) \\ \vdots \\ \psi(\nu_0) \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_s \end{pmatrix}$$

Ideas

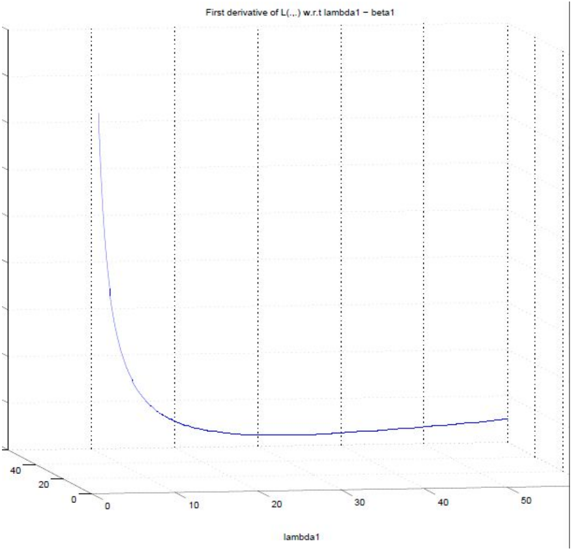
- recurrence formula for ψ:
  $\psi_i = \psi(1 + \sum_j \lambda_j g_j(x_i)) = \psi(\sum_j \lambda_j g_j(x_i)) + \frac{1}{\sum_j \lambda_j g_j(x_i)}$
- series expansions: not effective either
- asymptotic formulas: also not helpful
- done so far: numeric computation

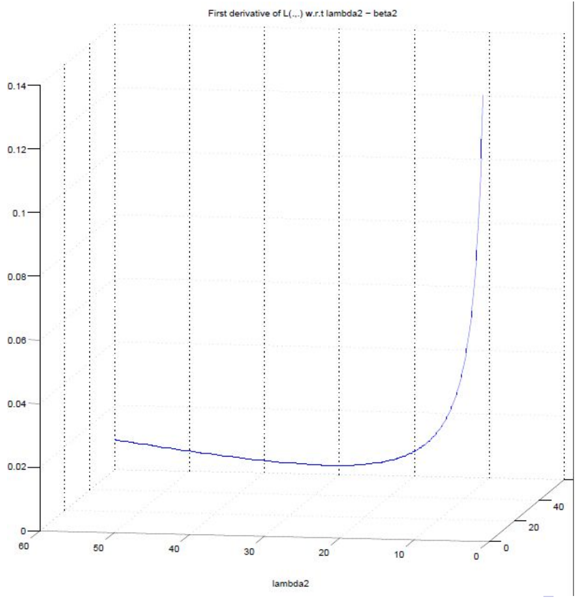
Example I

- number of realizations: 2
- 1 data source: $g_1 = (g_1(x_1), g_1(x_2)) = (0.3, 0.7)$
- results: $\lambda_1 = 40.25$, value: 0.0029
- $\hat{h}(x_1) = 0.31$, $\hat{h}(x_2) = 0.69$
- function converged
- 

Examples II

- 2 data sources: $g_1 = [0.3, 0.7]$, $g_2 = [0.4, 0.6]$
- results: $\lambda_1 = 56.4896$, $\lambda_2 = 53.0830$
- value: 0.0101, 0.0099
- $\hat{h} = (0.35, 0.65)$
- function converged


First derivative of L(.,.) w.r.t lambda1 − beta1
lambda1
0
10
20
30
40
50
40
20
0


First derivative of L(.,.) w.r.t lambda2 − beta2
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
60
50
40
30
20
10
0
40
20
0
lambda2

Examples III

- number of realizations: 2
- 5 data source: $\begin{pmatrix} 0.5 & 0.5 \\ 0.4 & 0.6 \\ 0.35 & 0.65 \\ 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$, $\lambda = \begin{pmatrix} 35.2 \\ 46.6 \\ 52.3 \\ 35.2 \\ 35.2 \end{pmatrix}$
- $\hat{h} = (0.44, 0.56)$
- function converged
- problem: if $g_j = (0.8, 0.2)$ is used in this example, algorithm does not converge