Non parametric estimation

Motivation: Posterior mean formula

$$t_G(Y) = E(\alpha|Y) = \frac{\int \alpha p(y|\alpha) dG(\alpha)}{\int p(y|\alpha) dG(\alpha)}$$

where the denominator $f(Y) = \int p(y|\alpha) dG(\alpha)$ is the marginal density of y.

We can either **directly** esimate $f(Y)$ and $f'(Y)$ or $estimate the $G(\alpha)$ 1. f-modelling: ignore the monoticity of $t_G(Y)$. can be improved by imposing monoticity/shape constraint in the kernel density estimation. 2. G-modelling: various ways estimate $G(\alpha)$, NPMLE, EM etc.

Computation

NPMLE and interior point method

Formulate the primal problem Following `KieferWolfowitz1956` we define our primal problem as

$$\max_G \{ \sum_i \log f(y_i) \}$$

$$\max_G \{ \sum_i \log(\int p(y_i|\alpha) dG(\alpha)) \}$$

As long as $p(y|\alpha)$ belongs to the exponential family, a solution G^* exists and is a discrete probability measure with no more than n atoms/mass points in the intervAL $(\min(Y), \max(Y))$. The solution is a highly parsimonious distribution.

The above problem is a convex optimization problem (rewrite as a minimization problem)

$$\min_G \left\{ -\sum_i \log g(y_i) \middle| g(y_i) = \int p(y_i|\alpha) dG(\alpha), \forall i \right\}$$

Since integral is a **linear** opertaor. The primal problem can be written explicitly as

$$\min_{f=dG} \left\{ -\sum_i \log g(y_i) \middle| g_i = T(f_i),\ K(f_i) = 1,\ \forall i \right\}$$

where

$$T(f_i) = \int p(y|\alpha) f_i d\alpha$$

and

$$K(f_i) = \int f_i d\alpha$$

Discretize We want to replace the linear operator by matrix multiplication (discretize the integral). Thus, the problem becomes

$$\min_{f=dG} \left\{ -\sum_i \log g(y_i) \middle| g = Af,\ 1^T f = 1 \right\}$$

where

$$A_{ij} = p(y_i|\alpha_j)$$

and

$$f = (f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_m))^T$$

where u_j are the grid points in the interval $(\min(Y), \max(Y))$.

Find the dual problem See appendix for the derivation:

$$\max_{\lambda,\mu}\left\{\sum_i \log\lambda_1(i)\middle| A^T\lambda_1 < \lambda_2 1,\ (\lambda_1 > 0)\right\}$$

Efrom log spline

See appendix.

EM algorithm

See appendix

Appendix

Dualization

Given the primal problem

$$\min_{f=dG}\left\{-\sum_i \log g(y_i)\middle| g = Af,\ 1^T f = 1,\ f > 0\right\}$$

The Lagragian is

$$L(f,\lambda,\mu) = -\sum_i \log g(y_i) + \lambda_1^T(g - Af) + \lambda_2(1^T f - 1) - \mu^T f$$

The primal objective function is $P(f) = \sup_{\lambda,\mu>0} L(f,\lambda,\mu)$. The dual objective function is $D(\lambda,\mu) = \inf_f L(f,\lambda,\mu)$.

Take derivative of L w.r.t. f

$$\frac{\partial L}{\partial f} = -A^T\lambda_1 + \lambda_2 1 - \mu = 0$$

Thus, $A^T\lambda_1 < \lambda_2 1$

Take derivative of L w.r.t. g_i

$$\frac{\partial L}{\partial g_i} = -\frac{1}{g_i} + \lambda_1(i) = 0$$

The dual problem can be written as

$$\max_{\lambda,\mu}\left\{\sum_i \log\lambda_1(i)\middle| A^T\lambda_1 < \lambda_2 1,\ (\lambda_1 > 0)\right\}$$

> PARFAIT!