Likelihood-based inference for probabilistic graphical models: Some preliminary results

Marco Cattaneo
Department of Statistics, LMU Munich
cattaneo@stat.uni-muenchen.de

PGM 2010, Helsinki, Finland
14 September 2010

general problem:

given:

- a set Θ of parameters θ corresponding to **probability distributions** P_θ
- the **likelihood function** *lik* on Θ induced by some data: $lik(\theta) \propto P_\theta(\text{data})$
- a function $g : \Theta \to \mathbb{R}^+$ describing the **quantity of interest** $g(\theta)$

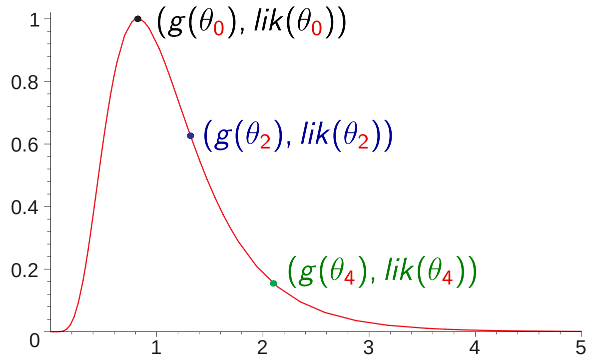
calculate: the **profile likelihood function** lik_g on $\mathbb{R}^+$ defined by

$$lik_g(x) = \max_{\substack{\theta \in \Theta : \\ g(\theta) = x}} lik(\theta) \qquad \text{for all } x \in \mathbb{R}^+$$

basic idea:

if $\theta = \theta_\alpha$ maximizes $[g(\theta)]^\alpha \, lik(\theta)$ over all $\theta \in \Theta$ for some $\alpha \in \mathbb{R}$,
then it also maximizes $lik(\theta)$ over all $\theta \in \Theta$ such that $g(\theta) = g(\theta_\alpha)$,
and therefore $(g(\theta_\alpha), lik(\theta_\alpha))$ is **a point in the graph** of lik_g

by varying α, **the whole graph** of lik_g may be obtained;
in particular, θ_0 is a ML estimate of θ,
and $g(\theta_\alpha)$ is an increasing function of α



example (naive classifier):

$$class \in \{a, b\}$$

$$color \in \{red, blue\} \qquad size \in \{BIG, small\}$$

complete training dataset:

class	#	#red	#BIG
a	48	7	39
b	52	12	30

goal: profile likelihood function for $P(class = a \mid color = red,\ size = BIG)$

problem formulation:

- each **naive Bayes classifier** is described by a parameter

$$\theta = \left(\theta_a, \theta_{red\,|\,a}, \theta_{red\,|\,b}, \theta_{BIG\,|\,a}, \theta_{BIG\,|\,b}\right) \in \Theta = [0,1]^5$$

- the above training dataset induce the **likelihood function** lik on Θ defined by

$$lik(\theta) \propto \theta_a^{48}\,(1-\theta_a)^{52}\,\theta_{red\,|\,a}^{7}\,(1-\theta_{red\,|\,a})^{41}\,\theta_{red\,|\,b}^{12}\,(1-\theta_{red\,|\,b})^{40}\,\theta_{BIG\,|\,a}^{39}\,(1-\theta_{BIG\,|\,a})^{9}\,\theta_{BIG\,|\,b}^{30}\,(1-\theta_{BIG\,|\,b})^{22}$$

- as **quantity of interest** choose

$$g(\theta) = \frac{P(class=a \mid color=red, size=BIG)}{P(class=b \mid color=red, size=BIG)} = \theta_a\,(1-\theta_a)^{-1}\,\theta_{red\,|\,a}\,\theta_{red\,|\,b}^{-1}\,\theta_{BIG\,|\,a}\,\theta_{BIG\,|\,b}^{-1}$$

application of the basic idea: for each $\alpha \in [-7, 12]$,

$$[g(\theta)]^{\alpha}\,lik(\theta) \propto \theta_a^{48+\alpha}\,(1-\theta_a)^{52-\alpha}\,\theta_{red\,|\,a}^{7+\alpha}\,(1-\theta_{red\,|\,a})^{41}\,\theta_{red\,|\,b}^{12-\alpha}\,(1-\theta_{red\,|\,b})^{40}\,\theta_{BIG\,|\,a}^{39+\alpha}\,(1-\theta_{BIG\,|\,a})^{9}\,\theta_{BIG\,|\,b}^{30-\alpha}\,(1-\theta_{BIG\,|\,b})^{22}$$

corresponds to the "likelihood function" induced by a **modified "training dataset"**, and is thus maximized by the "ML estimate"

$$\theta_\alpha = \left(\tfrac{48+\alpha}{100}, \tfrac{7+\alpha}{48+\alpha}, \tfrac{12-\alpha}{52-\alpha}, \tfrac{39+\alpha}{48+\alpha}, \tfrac{30-\alpha}{52-\alpha}\right)$$

since $\alpha \mapsto g(\theta_\alpha)$ maps $[-7, 12]$ in $[0, +\infty]$,

$$\{(g(\theta_\alpha),\ lik\,(\theta_\alpha)) : \alpha \in [-7, 12]\}$$

is a parametric expression for the graph of lik_g, and therefore

$$\left\{\left(\tfrac{g(\theta_\alpha)}{1+g(\theta_\alpha)},\ lik\,(\theta_\alpha)\right) : \alpha \in [-7, 12]\right\}$$

is a **parametric expression for the graph** of the profile likelihood function for $P(class = a \mid color = red,\ size = BIG)$