Theoretical case study: entropy minimization

Dr Paresh Date

Brunel University, UK

June 2016

Case study I

- Consider a discrete probability distribution, $\{\pi_i, x_i\}$, with $\sum_i \pi_i = 1$, $\pi_i > 0$. x_i are support points.
- This is our *trial* distribution. Suppose that some new information about the distribution becomes available, *e.g.* its mean needs to be $\bar{x}$.
- We need to find a new measure $\{\tilde{\pi}_i, x_i\}$, such that

$$\sum_i \tilde{\pi}_i = 1, \tilde{\pi}_i > 0, \sum_i \tilde{\pi}_i x_i = \bar{x} \qquad (1)$$

and $\tilde{\pi}_i$ is 'as close as possible' to π_i.

Case study II

- A measure of closeness typically used is *relative entropy*:

$$K_\pi(\tilde{\pi}) = \sum_{i=1}^{n} \tilde{\pi}_i \log\left(\frac{\tilde{\pi}_i}{\pi_i}\right). \qquad (2)$$

- Note that $K''_\pi(\tilde{\pi}) > 0$, so that relative entropy is convex in $\tilde{\pi}$.
- Minimizing $K_\pi(\tilde{\pi})$ with respect to constraints (1) is thus a convex problem.

Case study III

-

$$L = \sum_{i=1}^{n} \tilde{\pi}_i \log\frac{\tilde{\pi}_i}{\pi_i} - \sum_{i=1}^{n} \lambda_i \tilde{\pi}_i + \nu_1\left(\sum_{i=1}^{n} \tilde{\pi}_i - 1\right) + \nu_2\left(\sum_{i=1}^{n} \tilde{\pi}_i x_i - \bar{x}\right),$$

so that $(\nabla_L)_i = 1 + \log\frac{\tilde{\pi}_i}{\pi_i} - \lambda_i + \nu_1 + \nu_2 x_i$.

- KKT for optimal $(\tilde{\pi}_i^\star, \lambda_i^\star, \nu_1^\star, \nu_2^\star)$ gives $\lambda_i^\star = 0$ and

$$1 + \log\frac{\tilde{\pi}_i^\star}{\pi_i} + \nu_1^\star + \nu_2^\star x_i = 0, \text{ i.e.}$$

$$\tilde{\pi}_i^\star = \pi_i \exp(-\nu_2^\star x_i)\exp(-1-\nu_1^\star) = \frac{\pi_i \exp(-\nu_2^\star x_i)}{\sum_{i=1}^{n} \pi_i \exp(-\nu_2^\star x_i)},$$

- where the last equation uses the fact that $\tilde{\pi}_i$ add up to 1. Thus minimizing K_π wrt $\tilde{\pi}$ is equivalent to a **scalar**, unconstrained minimization in ν_2: $f(\nu_2) = (\sum_i \tilde{\pi}_i x_i - \bar{x}) = 0$.