Predicting the EQ-5D Preference Index using the SF-12 with Finite Mixture Models

Marcelo Coca Perraillon

University of Chicago

December 4, 2013

(… sections omitted …)

Finite mixture models

- Assume outcome is a mixture of distributions with some probability

$$f(y) = \sum_{j=1}^{k} \pi_j f_j(y|\mathbf{x}; \boldsymbol{\theta}_j)$$

- First application: Karl Pearson (1894), crab data

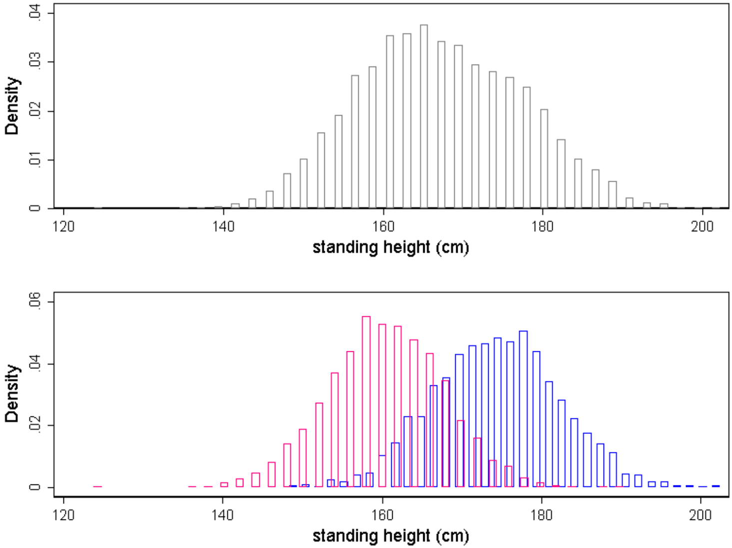
Example: Mixture of normals

- Mixture of two normal distributions

$$f(y_i|\boldsymbol{\theta},\boldsymbol{\pi}) = \sum_{j=1}^{2} \pi_j \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{1}{2\sigma_j^2}(y_i - \mathbf{x}_i'\boldsymbol{\beta}_j)^2\right)$$

- Objective is to estimate: σ_1, σ_2, π_1, and coefficients ($\boldsymbol{\beta}$)
- Model can be extended: $\pi_j(\mathbf{z}_i'\boldsymbol{\alpha}_j)$
- Distributions don't need to be of the same type
- Several of the key issues can be illustrated with a simple example

NHANES 2009-2010 data on height (combined and by sex, pink=women)



NHANES height mixtures

- Observed means: All: 167.2(10.3), Women: 160.5(7.32), Men: 174.4(7.93), $P_m = 0.48$
- This is not a problem that requires a mixture model. We do know the sex of NHANES responders. We're just pretending we don't.
- Fitted a mixture of two normals: $\hat{\beta}_0 w = 161.3$, $\hat{\sigma}_w = 7.6$, $\hat{\beta}_0 m = 175.3$, $\hat{\sigma}_m = 7.74$, $\hat{P}_m = 0.42$
- Two predictions: 1) Could calculate a weighted average: $\hat{\beta}_0 w \times (1 - \hat{P}_m) + \hat{\beta}_0 m \times \hat{P}_m$
- 2) Based on estimated parameters and observed height, classify observations and then predict conditional on class

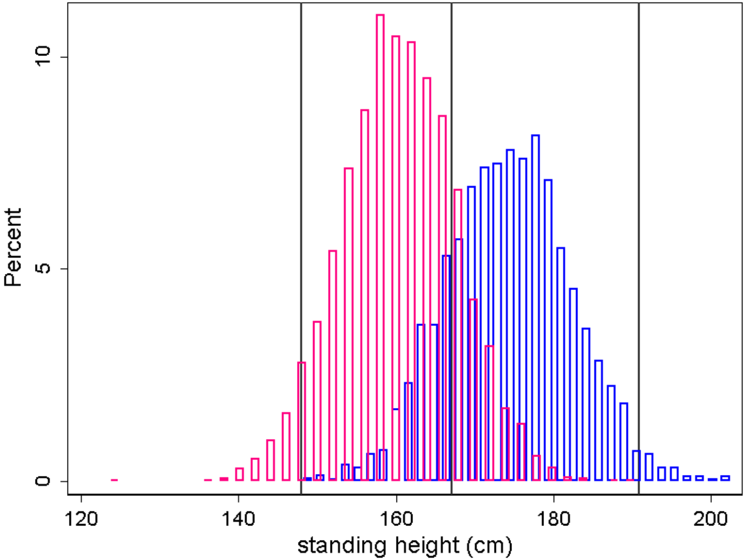
Classification

- Probability that an observation belong to class c is:

$$Pr\left(y_i \in class\ c|\mathbf{x}_i, y_i, \hat{\boldsymbol{\theta}}\right) = \frac{\hat{\pi}_c f_c(y_i|\mathbf{x}_i, \hat{\boldsymbol{\theta}}_c)}{\sum_{j=1}^{c} \hat{\pi}_j f_j(y_i|\mathbf{x}_i, \hat{\boldsymbol{\theta}}_j)}$$

- Model with no covariates correctly classifies 83 percent of observations
- Added one covariate in model for mixture, $\pi_j(\mathbf{z}_i'\alpha_j)$: a measure of body fat (women tend to have more body fat than men)
- Result: 90 percent of observations are correctly classified into males or females
- Which individuals are harder to classify? Those of average height. Their posterior probabilities are close to 0.5
- Which individuals are easier to classify? Those who are either very tall or very short

Middle bar: mean height of missclassified observations. No missclassifications left or right of other two bars



(... sections omitted ...)