From ICSDS @ Nice, France December, 2024

Shifts in distribution, covariates, prior probability, etc

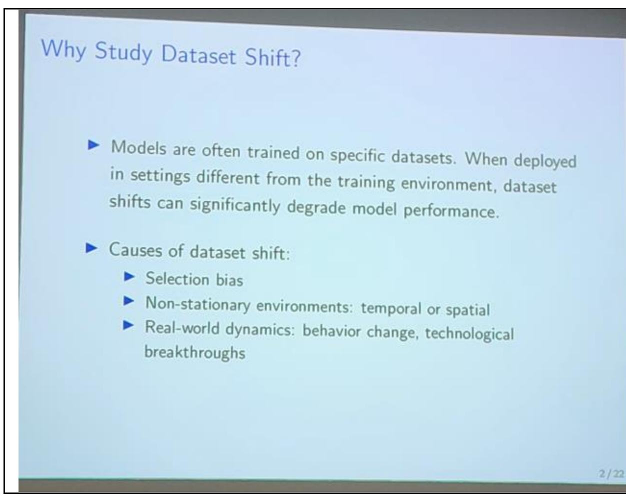
1. Huang MY, Qing J, Huang CY (2024) ``Efficient Data Integration Under Prior Probability Shift,” Biometrics, 2024, Mar 27, 80(2).
   https://pubmed.ncbi.nlm.nih.gov/38768225/
   https://profiles.ucsf.edu/chiung-yu.huang

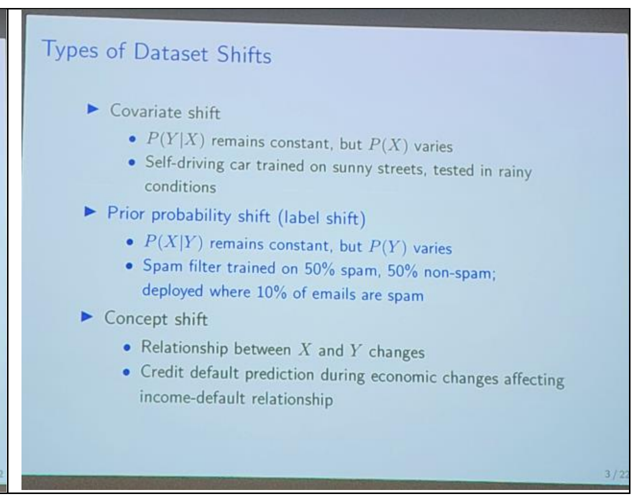
   Distribution-free prediction intervals under covariate shift, with an application to causal inference. Journal of the American Statistical Association. 2024. Qin J, Liu Y, Li M, **Huang C**Y.

2. Ying Jin, Naoki Egami, and Dominik Rothenhausler (2024) “Beyond Reweighting: On the Predictive Role of Covariate Shift in Effect Generalization”
   https://arxiv.org/pdf/2412.08869
   https://sites.google.com/view/rothenhaeusler/home

Some slides for the first talk:

Huang MY, Qing J, **Huang CY** (2024) ``Efficient Data Integration Under Prior Probability Shift,” Biometrics, 2024, Mar 27, 80(2).





General Dataset Shift Models

- Notation
  - $(Y_1, \mathbf{X}_1) \sim f_1(y, \mathbf{x})$ from training data
  - $(Y_2, \mathbf{X}_2) \sim f_2(y, \mathbf{x})$ from testing data
- A general dataset shift model

$$f_2(y, \mathbf{x}) = \frac{w(y, \mathbf{x}) f_1(y, \mathbf{x})}{\int\int w(u, v) f_1(u, v) du dv}$$

  or, equivalently, $f_2(y, \mathbf{x}) \propto w(y, \mathbf{x}) f_1(y, \mathbf{x})$
- $w(y, \mathbf{x})$ can be viewed as sampling weight function
  - Covariate shift: $w(y, \mathbf{x}) \equiv w(\mathbf{x}) \Rightarrow f_1(y \mid \mathbf{x}) = f_2(y \mid \mathbf{x})$
  - Prior probability shift: $w(y, \mathbf{x}) \equiv w(y) \Rightarrow f_1(\mathbf{x} \mid y) = f_2(\mathbf{x} \mid y)$
    $f_1(y \mid \mathbf{x}) \neq f_2(y \mid \mathbf{x})$
  - Concept shift: e.g. $w(y, \mathbf{x}) = w_1(y) w_2(\mathbf{x})$;
    $w(y, \mathbf{x}) = w(y, \mathbf{x}; \gamma)$

Data and Model Setup

- Two datasets
  $\mathcal{D}_1 = \{(\mathbf{X}_{1i}, Y_{1i}) : i = 1, \ldots, n_1\}$
  $\mathcal{D}_2 = \{(\mathbf{X}_{2i}, Y_{2i}) : i = 1, \ldots, n_2\}$
- Assume prior probability shift between $\mathcal{D}_1$ and $\mathcal{D}_2$:

$$f_1(\mathbf{x} \mid y) = f_2(\mathbf{x} \mid y)$$

- Assumed parametric model for $\mathcal{D}_1$: $f(y \mid \mathbf{x}; \theta)$
- Consider efficient estimation of θ under prior probability shift with $\mathcal{D}_1 \cup \mathcal{D}_2$

Likelihood Under Prior Probability Shift

- Suppose $\boldsymbol{X}_1 \sim dG_1(\mathbf{x})$ and $Y_2 \sim F_2(y)$
- Apply Bayes Rule and under prior probability shift

$$\begin{aligned} f_1(\mathbf{x} \mid y) &= \frac{f_1(y \mid \mathbf{x}; \boldsymbol{\theta}) dG_1(\mathbf{x})}{\int f_1(y \mid u; \boldsymbol{\theta}) dG_1(u)} \\ f_2(y, \mathbf{x}) &= f_2(\mathbf{x} \mid y) dF_2(y) \\ &= \frac{f_1(y \mid \mathbf{x}; \boldsymbol{\theta}) dG_1(\mathbf{x})}{\int f_1(y \mid u; \boldsymbol{\theta}) dG_1(u)} dF_2(y) \end{aligned}$$

Maximum Likelihood Estimation

- The MLE of F_2 is the empirical distribution, so $\widehat{q}_i = n_2^{-1}$.
- Profile likelihood: for each fixed θ, write

$$\widehat{\mathbf{p}}(\theta) = \operatorname*{argmax}_{\mathbf{p} \in \mathcal{P}_n} \ell(\theta, \mathbf{p}, \widehat{\mathbf{q}})$$

- Lagrange function:

$$\begin{aligned} \mathcal{L}(\mathbf{p}, \eta) = &\sum_{i=1}^{n} \log f(Y_i \mid \mathbf{X}_i; \theta) + \sum_{i=1}^{n} \log p_i \\ &- \sum_{i=1}^{n_2} \log \sum_{j=1}^{n} f(Y_{2i} \mid \mathbf{X}_j; \theta) p_j + \eta \left(\sum_{i=1}^{n} p_i - 1 \right), \end{aligned}$$

  η is the Lagrange multiplier.

Profile Likelihood Estimation

- For fixed θ, $\widehat{\mathbf{p}}(\theta)$ solves

$$p_i = \left\{ n_1 + \sum_{j=1}^{n_2} \frac{f(Y_{2j} \mid \mathbf{X}_i; \theta)}{\sum_{k=1}^{n} f(Y_{2j} \mid \mathbf{X}_k; \theta) p_k} \right\}^{-1}, \quad i = 1, \ldots, n.$$

- The equations can be rewritten as $\mathbf{p} = \mathbf{T}(\mathbf{p})$;
- Following a standard argument of fixed-point theory, the root can be obtained by iteratively computing $\mathbf{T}(\mathbf{p})$ with a proper initial point.

Prediction

Y_k^{new}: outcome corresponding to a new subject with $\mathbf{X}_k^{\text{new}}$

- Predict continuous response:

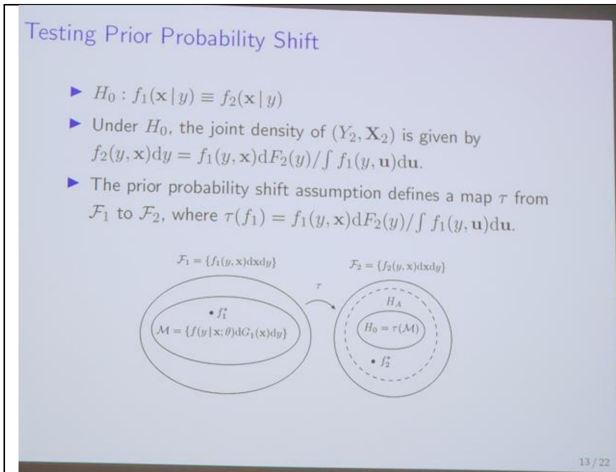
$$\widehat{Y}_1 = \int y f(y \mid \mathbf{X}_1^{\text{new}}; \widehat{\theta}) \mathrm{d}y \quad \text{and} \quad \widehat{Y}_2 = \frac{\int y f(y \mid \mathbf{X}_2^{\text{new}}; \widehat{\theta}) \mathrm{d}\widehat{R}(y)}{\int f(y \mid \mathbf{X}_2^{\text{new}}; \widehat{\theta}) \mathrm{d}\widehat{R}(y)},$$

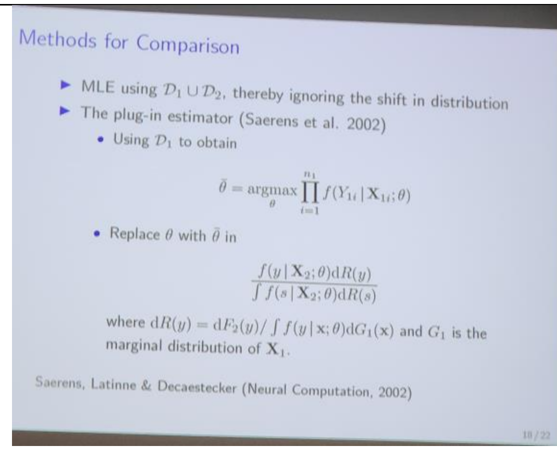
  where $\mathrm{d}\widehat{R}(y) = \mathrm{d}\widehat{F}_2(y) / \int f(y \mid \mathbf{x}; \widehat{\theta}) \mathrm{d}\widehat{G}_1(\mathbf{x})$.
- Predict discrete response:

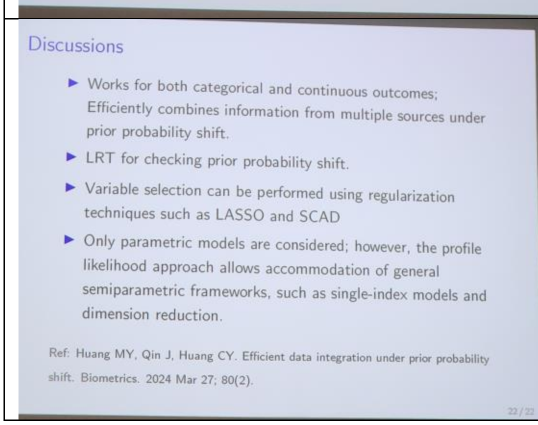
$$\widehat{Y}_1 = \operatorname*{argmax}_{y \in \{y_1, \ldots, y_K\}} f(y \mid \mathbf{X}_1^{\text{new}}; \widehat{\theta}) \quad \text{and}$$

$$\widehat{Y}_2 = \operatorname*{argmax}_{y \in \{y_1, \ldots, y_K\}} \frac{f(y \mid \mathbf{X}_2^{\text{new}}; \widehat{\theta}) \Delta \widehat{R}(y)}{\sum_{k=1}^{K} f(y_k \mid \mathbf{X}_2^{\text{new}}; \widehat{\theta}) \Delta \widehat{R}(y_k)},$$

  where $\Delta \widehat{R}(y) = \Delta \widehat{F}_2(y) / \int f(y \mid \mathbf{x}; \widehat{\theta}) \mathrm{d}\widehat{G}_1(\mathbf{x})$.







Some Slides from the 2nd talk