Killer Questions

- Can we identify a dataset containing the pediatric illness populations of interest and their immediate family members?
- Can we quantify the outcomes of interest and their relationship to the hypothesized mediating factors?
- Do we see any significant secondary effects of pediatric illness on family members, in comparison to a sensibly-defined control group?

Pediatric Illnesses

- Cancer
- Autism
- Asthma
- Cerebral Palsy
- Traumatic Event - being re-calculated

Family Member Outcomes

- Incidence of mental health diagnoses (depression or anxiety)
- Incidence of infection or illness requiring hospitalization
- Use of prescription drugs, specifically those for chronic pain
- Total cost of care

Next steps

- Look at the probability of a claim. Mixture distribution model on log scale given a claim, do you need health-care (yes/no),
- 2-part mixture model
- Specific mental health diagnoses
- Other predictors: median income, market segment, product description
- Other children as a covariate, indicator of other children.
- Look at young parents with young children, difference between adult and child.

CDF of a mixture model of a discrete probability at 0 and cdf F_T for values of $Y > 0$.

$$F(y) = \begin{cases} 0 & y < 0 \\ 1 - p & y = 0 \\ 1 - p + pF_T(y) & y > 0 \end{cases}$$

Define $\xi_i = \begin{cases} 1 & y_i > 0 \\ 0 & o.w. \end{cases}$ and $p = Pr(Y > 0)$

$$L(\theta|Y) = \prod_{i=1}^{n} p_i^{\xi}(1-p)^{1-\xi_i} f(y_i|\theta)^{\xi_i}$$

$$l(\theta|Y) = \sum_{i=1}^{n} \xi_i(\log p + \log f(y_i|\theta)) + (1-\xi_i)log(1-p)$$

`https://cran.r-project.org/web/packages/gmm/vignettes/gmm_with_R.pdf`
Generalized method of moments: Given a vector of functions of θ_0 with $E[g(\theta_0, x_i)] = 0$, there is often no solution to

$$\bar{g}(\theta) = \frac{1}{n}\sum_{i=1}^{n} g(\theta, x_i)$$

We can make it as close to zero as possible by minimizing the quadratic function $g(\theta)^T W g(\theta)$, where W is a positive definite and symmetric qxq matrix of weights. The optimal weight matrix is the inverse of the asymptotic variance $\Omega(\theta_0)$. The optimal matrix can be estimated by an heteroskedascity and autocorrelation consistent matrix (HAC) whose general form is

$$\hat{\Omega} = \sum_{s=-(n-1)}^{n-1} k_h(s)\hat{\Gamma}_s(\theta^*)$$

where $k_h(s)$ is a kernel and h is the bandwidth.

$$\hat{\Gamma}_s(\theta^*) = \frac{1}{n}\sum_i g(\theta^*, x_i) g(\theta^*, x_{i+s})^T$$

Original GMM algorithm Hansen (1982)

1. Compute $\theta^* = \arg\min \bar{g}(\theta)^T \bar{g}(\theta)$
2. Compute the HAC matrix $\hat{\Omega}(\theta^*)$
3. Compute the 2SGMM $\hat{\theta} \arg\min_\theta \bar{g}(\theta)^T [\hat{\Omega}(\theta^*)]^{-1} \bar{g}(\theta)$