STA 663 Final Project Outline

Joshua Fink

The paper I have chosen describes a type of finite mixture algorithm, one for modeling patterns of an outcome of interest as clusters across time. The method is termed by the authors as *Group-based Latent Trajectory Modeling*, and was developed by Daniel Nagin in 1999 (see attached paper).

The general specification for the binary version of the group based latent trajectory models is provided in equations (1) and (2). Once the researcher specifies the number of trajectories, j, the conditional distribution of the data for individual i, $y_{(i)}$, across his or her life is a function of time invariant covariates z_i and a time varying covariate x_i as specified in equation (1). The time-stable covariates, listed in the methods section as the conditioning variables (z_i), calculated the same across all models—the effect of these covariates on group membership is specified as a generalized logit function in equation (2). College enrollment (x_i), the time varying covariate in these models, affects the shape of trajectories and it is different for the count and binary data.

$$f(y_i, x_i) = \sum_{j}^{p} \mathrm{P}(V_i = j | Z_i = z_i) \mathrm{P}(Y_i = y_i | V_j, X_i = x_i) \tag{1}$$

$$\mathrm{P}(V_i = j | Z_i = z_i) = \frac{e^{\theta_k + \lambda_k z_i}}{\sum_{j}^{p} e^{\theta_k + \lambda_k z_i}} \tag{2}$$

This modeling technique can be adapted to Logistic—for binary outcomes, Poisson and zero-inflated Poisson—for count outcomes, and Gaussian—for continuous outcomes.

invariant covariates z_i and a time varying covariate x_i as specified in equation (1). The tim

covariates, listed in the methods section as the conditioning variables (z_i), calculated the s

acros   a ge

logit   ese r

affects the shape of trajectories and it is different for the count and binary data.

$$f(y_i, x_i) = \sum_{j}^{p} P(V_i = j | Z_i = z_i) P(Y_i = y_i | V_j, X_i = x_i)$$

$$P(V_i = j | Z_i = z_i) = \frac{e^{\theta_k + \lambda_k z_i}}{\sum_{j}^{p} e^{\theta_k + \lambda_k z_i}}$$

This modeling technique can be adapted to Logistic—for binary outcomes, Poisson and ze

Poisson—for count outcomes, and Gaussian—for continuous outcomes.