Glossary

categorical latent variable. A categorical latent variable has levels that represent unobserved groups in the population. Latent classes are identified with the levels of the categorical latent variables and may represent healthy and unhealthy individuals, consumers with different buying preferences, or different motivations for delinquent behavior.

class model. A class model is a regression model that is applied to one component in a mixture model. In the absence of covariates, the regression model reduces to a distribution function.

Class model is also referred to in the literature as a "component model", "component density", or "component distribution".

class probability. In the context of FMM, the probability of belonging to a given class. `fmm` uses multinomial logistic regression to model class probabilities.

Class probability is also referred to in the literature as a "latent class probability", "component probability", "mixture component probability", "mixing probability", "mixing proportion", "mixing weight", or "mixture probability".

EM algorithm. See *expectation-maximization algorithm*.

expectation-maximization algorithm. In the context of FMM, an iterative procedure for refining starting values before maximizing the likelihood. The EM algorithm uses the complete-data likelihood as if we have observed values for the latent class indicator variable.

finite mixture model. A finite mixture model (FMM) is a statistical model that assumes the presence of unobserved groups, called latent classes, within an overall population. Each latent class can be fit with its own regression model, which may have a linear or generalized linear response function. We can compare models with differing numbers of latent classes and different sets of constraints on parameters to determine the best fitting model. For a given model, we can compare parameter estimates across classes. We can estimate the proportion of the population in each latent class, and we can predict the probabilities that the observations in our sample belong to each latent class.

FMM. See *finite mixture model*.

generalized linear response functions. Generalized linear response functions include linear functions and include functions such as probit, logit, multinomial logit, ordered probit, ordered logit, Poisson, and more.

These generalized linear functions are described by a link function $g(\cdot)$ and statistical distribution F. The link function $g(\cdot)$ specifies how the response variable y_i is related to a linear equation of the explanatory variables, $\mathbf{x}_i\boldsymbol{\beta}$, and the family F specifies the distribution of y_i:

$$g\{E(y_i)\} = \mathbf{x}_i\boldsymbol{\beta} \qquad y_i \sim F$$

If we specify that $g(\cdot)$ is the identity function and F is the Gaussian (normal) distribution, then we have linear regression. If we specify that $g(\cdot)$ is the logit function and F the Bernoulli distribution, then we have logit (logistic) regression.

In this generalized linear structure, the family may be Gaussian, gamma, Bernoulli, binomial, Poisson, negative binomial, ordinal, or multinomial. The link function may be the identity, log, logit, probit, or complementary log–log.

latent class. A latent class is an unobserved group identified by a level of a categorical latent variable.

Latent class is also referred to in the literature as a “class”, “group”, “type”, or “mixture component”.

latent variable. See *categorical latent variable*.

pointmass density. In the context of FMM, a degenerate distribution that takes on a single integer value with probability one. A pointmass density is used in combination with other FMM distributions to model, most commonly, zero-inflated outcomes.

For suggested citations, see the FAQ on citing Stata documentation.