Hurdle Models

In standard count models, zeros and the nonzeros (positives) are assumed to come from the same generating process. With hurdle models, these two processes are not assumed to be the same. A binomial random variable determines the binary outcome of whether a count variate is zero or positive. If positive, the hurdle is crossed, and the distribution of the positives is a count data model.

With zero-inflated models, the response variable is modeled as a combination of a binomial random variable and a Poisson or negative binomial distribution.

A Hurdle model example would be a buy decision. First, you decide to buy then you decide on the quantity.

Hurdle models were first proposed by a Canadian statistician Cragg (1971), and later developed further by Mullahy (1986).

Cragg, J. G. "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods", Econometrica, 1971, 39, 829-844

Mullahy, J. "Specification and Testing of Some Modified Count Data Models", Journal of Econometrics, 1986, 33, 341-365

In R, hurdle count data models can be fitted with the hurdle function from the pscl package. Python does not have a hurdle package.