Discrete and Bayesian Choice Modeling for Consumer Preferences

July 29, 2025

Abstract

This document provides a synopsis of two choice modeling approachesDiscrete Choice Model (DCM) using logistic regression and Bayesian Choice Modelapplied to analyze consumer preferences from choice data. The DCM estimates utility coefficients deterministically, while the Bayesian model incorporates uncertainty, producing outputs like utilities with credible intervals and willingness-to-pay (WTP). A simulator based on the DCM predicts market shares under various price scenarios.

1 Discrete Choice Model (Logistic Regression)

The Discrete Choice Model (DCM) uses logistic regression to analyze consumer preferences from choice data stored in Excel files (`profiles.xlsx`, `data_final.xlsx`, `groups_final.xlsx`). Implemented in Python with `statsmodels`, it estimates utility coefficients for features such as price, size performance (e.g., high performance at 66kA), and advanced features (e.g., electrical life, health indication). The model encodes categorical variables via dummy coding, handles interactions (e.g., Price × Panel Builder), and uses SHAP values to assess feature importance. Outputs include utility plots (`utilities.png`), feature importance visualizations (`feature_importance_dcm.png`), and price elasticity analyses (`price_elasticity.png`).

1.1 Model Formulation

The utility of profile i for respondent n in choice set t, U_{nit}, is modeled as:

$$U_{nit} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{nik} + \epsilon_{nit}, \tag{1}$$

where β_0 is the intercept, β_k are coefficients for features x_{nik} (e.g., price, size performance), and ϵ_{nit} is the error term. The probability of choosing profile i is given by the logistic function:

$$P_{nit} = \frac{\exp(U_{nit})}{\sum_{j \in C_t} \exp(U_{njt})}, \tag{2}$$

where C_t is the choice set. The model is fitted using maximum likelihood estimation, with coefficients saved to `utilities_[group].xlsx` for group-specific analyses.

2 Bayesian Choice Model

The Bayesian Choice Model, implemented with `pymc`, adopts a probabilistic approach to model consumer choices, incorporating uncertainty via 95% credible intervals. It processes the same Excel data, encoding features and fitting a categorical choice model with a softmax link function. The model estimates utility coefficients (`betas`) and willingness-to-pay (WTP) for features, producing key outputs:

- **Utilities with Uncertainty**: A forest plot (`utilities_with_uncertainty_bdcm.png`) visualizes feature utilities with 95% credible intervals, saved as `beta_coefficients.xlsx`.
- **Willingness-to-Pay**: A horizontal bar plot (`wtp_analysis_enhanced_bdcm.png`) shows WTP estimates with 95% highest density intervals (HDI), saved as `wtp_results_bdcm.xls`

Additional outputs include market share predictions and price scenario simulations (`price_scenario_sh`

2.1 Model Formulation

The utility for profile i, respondent n, and choice set t is:

$$U_{nit} = \sum_{k=1}^{K} \beta_k x_{nik}, \tag{3}$$

where $\beta_k \sim \text{Normal}(0, 5)$ are feature coefficients with a prior distribution, and x_{nik} are feature values. The choice probability uses the softmax function:

$$P_{nit} = \frac{\exp(U_{nit})}{\sum_{j \in C_t} \exp(U_{njt})}. \tag{4}$$

WTP for feature k is calculated as:

$$\text{WTP}_k = -\frac{\beta_k}{\beta_{\text{Price}} + \epsilon}, \tag{5}$$

where β_{Price} is the price coefficient, and $\epsilon = 10^{-5}$ prevents division by zero. The model is fitted using MCMC, with posterior distributions visualized for diagnostics (`convergence_diagnostics.` `posterior_predictive_check.png`).

3 Simulator

The simulator, built on the DCM logistic regression model, predicts market shares for product profiles under price scenarios (e.g., baseline, ±20% price changes, custom prices). It computes choice probabilities using the fitted model:

$$P_i = \frac{\exp(\beta_0 + \sum_{k=1}^{K} \beta_k x_{ik})}{\sum_{j=1}^{J} \exp(\beta_0 + \sum_{k=1}^{K} \beta_k x_{jk})}, \tag{6}$$

where P_i is the market share for profile i, and J is the number of profiles. Outputs include visualizations (`profile_shares_line.png`) and Excel files (`Scenarios.xlsx`). The Bayesian models outputs enhance interpretation but are not directly used in the simulator.