Sawtooth Software

RESEARCH PAPER SERIES

Bayesian Smoothing in HB-MNL: An Intuitive Explanation

Bryan Orme
Sawtooth Software

3210 N. Canyon Rd., Provo, Utah
+1 801 477 4700
www.sawtoothsoftware.com

Bayesian Smoothing in HB-MNL: An Intuitive Explanation

Bryan Orme, Sawtooth Software

July, 2026

Introduction

With HB estimation of utilities for CBC or MaxDiff, the aim is not to maximize the fit to each respondent's choices. Choice questionnaires, especially CBC, are typically sparse at the individual respondent level. Even the best respondents are human and can be at least a little inconsistent (they answer with some noise). Choice tasks are realistic and easy for respondents, but each choice provides only limited information. We observe which alternative was chosen, but not whether it was only slightly preferred or strongly preferred to the competing alternatives.

Maximum likelihood individual-level utilities would capitalize on noise and overfit, often (especially in the case of CBC) leading to poor predictions of consumer behavior. Such utilities often have many reversals—out of order attribute levels that run counter to logical relationships (such as a low quality level preferred to a high quality level).

To avoid problems stemming from sparse data and overfitting at the individual level, HB finds an appropriate balance between fitting each respondent's choices well and being smoothed toward ("borrowing" information from) the population preferences. This short article explains how that happens, using an intuitive approach and a tiny bit of math.

When Respondents Are Not Very Consistent

In Sawtooth's training slides for our CBC (Choice-Based Conjoint) course, in the section where we introduce HB estimation of utility scores, we state:

- **If the respondent's choices are poorly fit, then their estimated utilities depend more on the population distribution, and are influenced less by their own individual data.**

I thought it might be helpful to explain more fully how that happens. After all, it's a very good aspect of HB estimation—it's like an automatic noise reduction procedure. Respondents whose choices provide little reliable information receive estimates that rely more heavily on the population distribution than on their own data.

In Sawtooth's HB-MNL algorithm (hierarchical Bayes Multinomial Logistic regression), we employ the common Metropolis-Hastings Algorithm to find individual-level utilities that

both fit the respondent's choices well and also have a high likelihood of coming from the population distribution of preferences.

Here's how it works:

In each iteration of the HB algorithm, it makes a new proposal (a set of utility scores) and evaluates whether that proposal is good or bad. That evaluation multiplies the **Likelihood** that the utilities explain the respondent's choices by the likelihood that the utilities would come from the population distribution (the **Prior Density**).

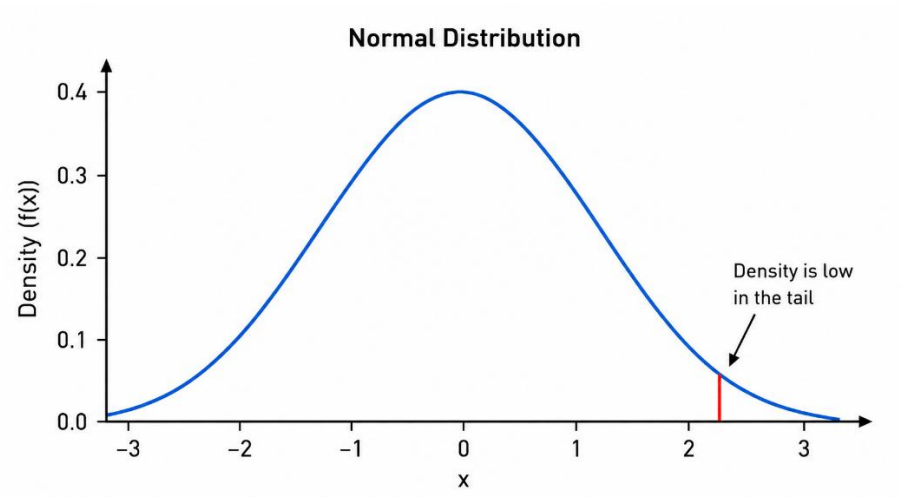
Posterior Probability ∝ Likelihood x Prior Density

(where ∝ means "is proportional to").

You can think of Likelihood and Prior Density as two competing influences. The Likelihood rewards utility values that explain the respondent's own choices well, while the Density rewards utility values that are typical of the overall population.

The **Likelihood** of how well a set of utilities fits the respondent's data is pretty easy to understand. For each choice task, you use the estimated utility scores to calculate the utility of each product concept. Those values are then converted into probabilities (using the logistic equation) that each concept would be chosen. The model then looks at the probability assigned to the concept the respondent actually selected. If the model consistently assigns high probabilities to the respondent's actual choices across all the choice tasks, then the likelihood is high.

Regarding the **Prior Density**, imagine a normal population distribution where a utility proposal ends up in the skinny tail. The height of the curve (the Density) in the tail is much lower than the height in the fat part of the distribution.



Of course, with a conjoint design with multiple attributes and multiple levels per attribute, we're dealing with a multivariate distribution in multidimensional space, but it's the same idea.

Now, here's where the scrubbing (Bayesian smoothing) magic happens for HB-MNL when a respondent's choices are noisy. Many different utility vectors fit the observed data almost equally well (or poorly, in this case). Remember that the likelihood that we'll accept a new proposal in HB-MNL for that respondent's utilities is proportional to:

Likelihood x Prior Density

So, most any proposal for the respondent's utilities will have approximately equal Likelihood. If there are four concepts in the questionnaire, the Likelihood for each choice task will be right around the chance level, ¼ = 0.25.

Since just about any proposal for the respondent's utilities leads to similar Likelihood, the **Prior Density** part of the equation becomes the dominant influence, and proposals that are closer to the fatter part of the population distribution will be the ones that tend to be accepted. Unusual proposals that don't look very likely to come from the population's preferences are highly discouraged.

When There Are Numerous Choices per Respondent

In another part of Sawtooth's CBC training, we emphasize:

- **The more choices a respondent makes, the more the lower level model (the respondent's data) influences their final utilities. Less smoothing to the population means occurs.**

Back when I was first learning about HB-MNL, Sawtooth's founder Rich Johnson told me why this occurs. He explained it as follows: with more choice tasks, the likelihood is formed by multiplying more and more choice probabilities. Imagine a respondent who completes 20 choice tasks. If a utility proposal assigns a probability of 75% to each observed choice (3x the chance level for choices among quads, which is quite good), the overall likelihood still becomes quite small: $0.75^{20} = 0.003171$. More importantly, even modest differences in how well two utility proposals fit the respondent's choices compound across tasks. As a result, the ratio between their likelihoods grows larger and larger. The Likelihood therefore increasingly dominates the Density when deciding which proposals are accepted.

So, to the degree that respondents a) answer consistently, and b) complete many choice tasks, there is less Bayesian smoothing to the upper-level model (the population means and covariances).

Summary and Conclusion

Wrapping it up, HB-MNL makes excellent use of the data for CBC and MaxDiff modeling. If respondents give us inconsistent answers, we limit their negative effects on predictions by smoothing them toward the population mean. On the other hand, if we've got a lot of information (choice tasks) about a respondent and they tend to be fairly consistent, we do relatively less smoothing toward the population mean—and that little bit of smoothing typically improves our representation of that respondent's preferences, given the sparse nature of most choice experiments. This also means that if you're hoping to bring out greater differences between respondents and segments of respondents (greater heterogeneity), you'll not want to use just a few choice tasks (such as 5 or 6) in CBC and you'll want to show each item at least 2x per respondent in MaxDiff.