The HE

How Bayesian methods have changed the face of marketing research.

Reprinted with permission from *Marketing Research*, Summer 2004, published by the American Marketing Association.

B revolution

For most of the history of marketing research, analytical methods have relied on "classical" statistics. We use these methods to make inferences about the characteristics of a population from the characteristics of a sample drawn from that population. Because we can never determine the true value of the population parameters from samples with complete certainty, we have regarded the sample as one realization of many that could have been realized and then quantify our uncertainty with a frequency concept of probability. If we're interested in the population mean for some measure (say height, for example), then the sample mean is our best guess about the mean of the population, and we conceive of uncertainty as arising from hypothetical samples that could have been selected from the population. Classical statistical methods express uncertainty through the variability of statistics calculated from these hypothetical samples when interpreting confidence intervals and conducting hypothesis tests.

Over the last 10 years, a paradigm shift has occurred in the statistical analysis of marketing data, especially data obtained from conjoint experiments. The new paradigm reflects a different perspective on probability. This view of probability was first proposed in the 18th century by Thomas Bayes, an English clergyman who died in 1760. Bayes wrote a paper (published posthumously in 1763) in which he proposed a rule for accounting for uncertainty that has become known as Bayes' Theorem. This became the foundation for Bayesian statistical inference, and Bayes most likely would be amazed by the many applications sprouting from his paper. In Bayesian statistics, probabilities reflect a belief about the sample of data under study rather than about the frequency of events across hypothetical samples. Bayes' Theorem exploits the fact that the joint probability of two events, A and B, can be written as the product of the probability of one event and the conditional probability of the second event, given the occurrence

By Greg M. Allenby, David G. Bakken, and Peter E. Rossi

Executive Summary

In the last 10 years, a paradigm shift has occurred in the statistical analysis of marketing data. Since the early 1990s, more than 50 papers on hierarchical Bayes (HB) methods have been published in top marketing journals, offering better solutions to a wider class of research problems than previously possible. This article provides an introduction to and perspective on Bayesian methods in marketing research.

of the first event. If we consider "A" to be our "hypothesis" (H) in the absence of any observations "B" (or "D" for data), we can express the rule as follows:

$$\Pr(H|D) = \Pr(D|H) \times \Pr(H) / \Pr(D)$$

The probability of the hypothesis given (or "conditional" on—the meaning of the "|" symbol) the data is equal to the probability of the data given the hypothesis times the probability of the hypothesis divided by the probability of the data. We refer to the Pr(H) as the prior probability and Pr(H|D) as the posterior probability. (See "Bayes' Theorem Example" on page 25.) While prior probabilities can be thought of as arising from hypothetical samples, the frequency concept of probability need not be employed, and the prior probability can reflect the researcher's belief about the state of nature. Alternatively, "H" can stand for any unobserved aspect of an analysis, including model parameters such as the estimates of utility in a conjoint model. In Bayesian analysis, all unobserved objects of analysis are dealt with in the same manner.

In Bayesian statistics, probabilities reflect a belief about the sample of data under study rather than about the frequency of events across hypothetical samples.

One of the main differences between the Bayesian and the classical approach is that the classical approach assumes that the hypothesis (H) is known with certainty, and the methods search for the best fitting model that maximizes Pr(D|H)—the probability of the data given the hypothesis. In contrast, the Bayesian approach acknowledges that the hypothesis, in reality, is never known, and the classical assumption that it is known limits the ability of the analyst to account for uncertainty. (The data may be consistent with many hypotheses.) The Bayesian approach asks, "Which hypothesis is most likely, given the data?" It is the observed data that are known for certain, and Bayesian analysis then proceeds by "conditioning" on the data.

Bayesian Application Emergence

Until recently, Bayesian statistical inference has been more a subject of intellectual curiosity than a practical application among researchers in marketing and other fields. There are at least three reasons for this. First, for many simple problems where the analyst is unwilling or unable to make informative statements about prior probabilities, the Bayesian and classical approach lead to essentially the same conclusions. Second, while Bayes' Theorem is a conceptually simple method of accounting for uncertainty, it has been difficult to implement in all but the simplest problems. As an example, a conjoint analysis involving 15 part-worth estimates and 500 respondents leads to an analysis with 750 parameters, making the application of Bayes' Theorem difficult. Finally, Bayesian analysis is particularly well-suited for understanding what we might call the "data generating process" behind complex marketing behaviors. Marketing researchers, however, do not usually make inferences about the process that gives rise to the observed behavior. The one area in which we do pay attention is the process by which consumers maximize utility in making choices. It's not surprising, therefore, that Bayesian analysis has gained a strong following in this area.

Most marketing researchers are familiar with a least one of the variations of conjoint analysis (full profile conjoint, adaptive conjoint, and "choice-based" conjoint). In conjoint analysis, consumers are presented with descriptions of actual or hypothetical products or services and asked to evaluate these products or services in some fashion. In the language of modeling, the characteristics that describe the alternatives (attributes or features, for example) and the consumer behavior (choice among alternatives, for example) are observed. The utilities that consumers assign to the different product characteristics are unobserved. Random factors (e.g., response errors) are also unobserved. The analytical challenge lies in isolating the unobserved utility from the other unobserved factors, given the attribute descriptions and observed choices. Moreover, it is desirable to do this in a manner that allows for respondent heterogeneity. A shortcoming of classical methods for analyzing conjoint data is that they cannot easily account for heterogeneity across consumers in the unobserved utilities.

Accounting for uncertainty is important when analyzing marketing research data. Marketing data typically comprises many heterogeneous "units" (e.g., households, respondents, panel members, activity occasions) with limited information on each unit and large amounts of uncertainty. In a conjoint analysis, it is rare to have more than 20 or so evaluations or choices of product descriptions per respondent. In direct marketing data, it is rare to have more than a few dozen orders for a customer within a given product category. Researchers often report that predictions in marketing research analyses are too aggressive and unrealistic. A reduction in price of 10% or 20%, for example, results in an increase in market share that's known to be too large. Overly optimistic predictions can easily result if estimates of price sensitivity are incorrectly assumed to be known with certainty. When uncertainty is taken into consideration, market share predictions become more realistic and tend to agree with current and past experience. Bayes' Theorem therefore offers an elegant solution to an important problem.

HB Models

Recent developments in statistical computing have made Bayesian analysis accessible to researchers in marketing and other fields. A key innovation from the 1980s, known as Markov Chain Monte Carlo (MCMC) simulation, has facilitated the estimation of complex models of behavior that can be infeasible to estimate with alternative methods. (Monte Carlo simulators provide a mechanism for generating random draws from statistical distributions. The Markov Chain is a device that, coupled with the Monte Carlo simulator, transforms the simulator into a rather efficient engine for searching the randomly generated distribution.) These models are written in a hierarchical form and, when estimated with MCMC methods, are referred to as hierarchical Bayes (HB) models. Discrete choice models, for example, assume that revealed choices reflect an underlying process where consumers have preferences for alternatives and select the one that offers greatest utility. Utility is assumed to be related to specific attribute levels that are valued by the consumer, and consumers are assumed to be heterogeneous in their preference for the attributes. The model is written as a series of hierarchical algebraic statements, where model parameters in one level of the hierarchy are "unpacked," or explained, in subsequent levels:

(1) $\Pr (y_{ih} = 1) = \Pr (V_{ih} + \varepsilon_{ih} > V_{jh} + \varepsilon_{jh}$ for all j)

(2) $V_{ih} = x_i'\beta_h$

(3) $\beta_h \sim \text{Normal}(\bar{\beta}, \Sigma_\beta)$

Here "Pr" stands for "the probability that…," "i" and "j" denote different choice alternatives, y_{ih} is the choice outcome for respondent h, V_{ih} is the utility of choice alternative i to respondent h, and ε_{ih} is a random or "error" component. In equation 2, x_i denotes the attributes of the i^{th} alternative, and β_h are the weights given to the attributes by respondent h. Finally, equation 3 is a "random-effects" model that assumes that the attribute weights are normally distributed in the population (with a mean of and $\bar{\beta}$ variance of Σ_β).

The "bottom" of the hierarchy specified by equations 1–3 is the model for the observed choice data. Equation 1 represents the individual's choice process and specifies that alternative j is chosen if the latent or unobserved utility is the largest among all of the alternatives. Latent utility is not observed directly and is linked to characteristics of the choice alternative in equation 2. Each individual's partworths or attribute weights (the bh in equation 2) are linked to those of other individuals by a common distribution in equation (3). Equation 3 allows for heterogeneity among the units of analysis by specifying a probabilistic model of how the units are related. Bayes' Theorem therefore provides a method of "bridging" the analysis across respondents while providing an exact accounting of all the uncertainty present. Even casual observation of consumer choices suggests that consumers vary in the value they assign to different product features or attributes. The advantage of the Bayesian approach to this problem is the ability to reflect the heterogeneity in those consumer preferences.

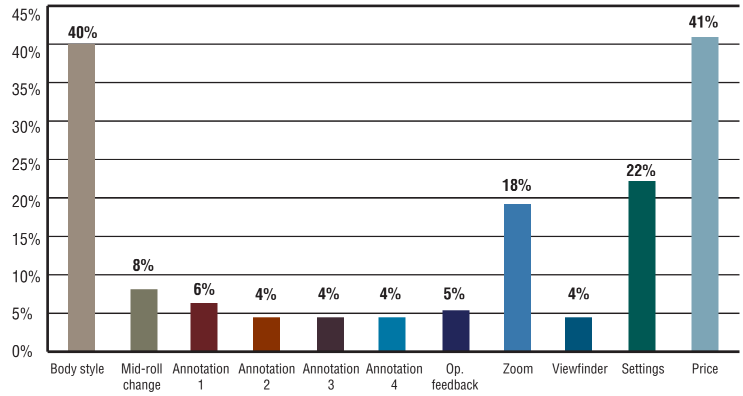
One of the main objections to choice-based conjoint has been that the method enables us to estimate utilities only in the aggregate, averaged across all respondents in the survey. The HB method, in contrast, allows us to make "individual-level" estimates of utilities for each feature level. This seemingly overcomes this objection, making choice-based conjoint more comparable to ratings or ranking-based methods of analysis. However, the individual level utilities estimated by OLS regression for ratings and rank data are not particularly reliable, given the small number of observations for any single individual. A principal benefit of HB estimation lies in equation 3, which bridges analysis among respondents in the sample and increases the information used to make inference about each respondent's utilities.

The predictive superiority of HB methods stems from the freedom afforded by MCMC to specify more realistic models and the ability to conduct disaggregate analysis. Models of heterogeneity were previously limited to the use of demographic covariates to explain differences or the use of finite mixture models, such as latent class models. Neither approach is realistic—demographic variables are too broad in scope to be related to attributes in a specific product category, and the assumption that heterogeneity is well-approximated by a small number of customer segments or "latent classes" is more a hope than a reality. Much of the predictive superiority of HB methods is due to avoiding the restrictive analytic assumptions that alternative methods impose.

Solving Marketing Problems

The best way to illustrate the power of Bayesian methods to solve marketing problems is with a concrete example. Most practitioners of conjoint methods have observed, at least anecdotally, that consumers may reject some feature levels outright. Put differently, these feature levels do not enter into the consumer's consideration set. For example, some consumers may simply reject all choices that exceed a certain price threshold. The basic conjoint model assumes, however, that consumers have at least some minimal utility for all feature levels and, moreover, that they apply a "compensatory rule" in making choices or evaluating alternatives. The compensatory rule implies that consumers will consciously trade off a better level of one feature in order to get a more desirable level of a second feature. In reality, consumers may use non-compensatory rules in making their choices. One such rule is the "conjunctive" or "and" rule, which requires that an alternative have certain features in order to be considered.

Exhibit 1 Respondents screening on camera attributes



Practitioners have tried different ad hoc or heuristic approaches for reducing the bias that departures from the assumption of a compensatory decision process might introduce into the estimation of conjoint utilities. One approach involves asking respondents to eliminate feature levels that are unacceptable to them. These features may then be eliminated from the conjoint experiment for that respondent (possible only in ACA or adaptive conjoint) or the information might be used in estimating the utility values, such that the unacceptable levels are assigned a large negative utility.

The real problem for the modeler is that, while the choice behavior is observed, the formation of the consideration set or the decision rule is not. The question becomes, can we infer the consideration set and the decision rule from the observed choices? Happily, the answer is "yes." Tim Gilbride, a graduate student at Ohio State University, working with Greg Allenby (one of the authors of this article), has developed a modeling approach that uses Bayesian estimation to simultaneously estimate the attribute-level utilities and test for different consideration sets and decision rules. Using a choice-based conjoint study for digital cameras, Gilbride estimated a variety of models, including a "standard" HB choice model, a model that assumed a conjunctive ("and") rule, a model that assumed a disjunctive ("or") rule, and a model that allowed for heterogeneity in the decision process (some consumers use compensatory rules, some use conjunctive rules, etc.). The key finding is that most consumers, at least in this category, appear to use a conjunctive rule. Moreover, the model makes it possible to identify the frequency with which respondents screen on each of the attributes. Exhibit 1 shows the results of this analysis for the digital camera study.

The Challenges

Freedom from computational constraints allows researchers and practitioners to develop more realistic models of buyer behavior and decision making. Moreover, this freedom enables exploration of marketing problems that have proven elusive over the years, such as models for advertising ROI, sales force effectiveness, and similarly complex problems that often involve endogeneity. This occurs when the dependent variable

Bayes' Theorem Example

A person working as a quality inspector on a production line is tasked with picking out units that have a specific blemish. Testing has determined that this particular inspector correctly declares (D) 90% of the units that have blemishes (B+), and never incorrectly declares an unblemished unit (B-). The inspector picks up a unit from the production line and declares it to be blemished. What is the probability that this unit is in fact blemished? We might be inclined to guess that the probability is 90%, but we need to take account of the fact that the inspector never makes a mistake if the unit is unblemished. We can use Bayes' Theorem to calculate the odds that the unit in the inspector's hand is in fact blemished. From Bayes' Theorem we have:

$$\Pr(B+|D) = \Pr(D|B+) \times \Pr(B+) / \Pr(D)$$

$$\Pr(B-|D) = \Pr(D|B-) \times \Pr(B-) / \Pr(D)$$

Dividing the two expressions to remove Pr(D) leaves:

$$\frac{\Pr(B+|D)}{\Pr(B-|D)} = \frac{\Pr(D|B+)}{\Pr(D|B-)} \times \frac{\Pr(B+)}{\Pr(B-)}$$

or "posterior odds = likelihood ratio × prior odds." If a blemish is present, the inspector is correct 90% of the time, so $\Pr(D|B+) = 0.90$. When a blemish is not present, the inspector never makes a mistake, so $\Pr(D|B-) = 0.00$. This means that, regardless of the prior odds that a blemish exists, $\Pr(B+)/\Pr(B-)$. Once the inspector declares a blemish, the posterior odds are infinite or the probability that a blemish exists, $\Pr(B+|D)$, is equal to one and the probability that a blemish does not exist, $\Pr(B-|D)$ is zero. For the less extreme case where the inspector occasionally makes a mistake when a blemish is not present, Bayes' Theorem accounts for all forms of uncertainty in deriving the posterior odds.

has some impact on the independent variable as, for example, when advertising budgets are established as a percent of the previous year's sales. The promise of Bayesian statistical methods lies in the ability to deal with these complex problems, but the very complexity of the problems creates a significant challenge to both researchers and practitioners.

In most cases, there exists no off-the shelf programs for estimating complex HB models. Sawtooth Software has created two programs for estimating HB models, one for choice-based conjoint and one for OLS regression. However, these programs are limited to modeling relatively simple problems. In fact, these programs are really HB extensions to "standard" choice-based conjoint and regression models. While the availability of these programs from Sawtooth Software has been a major impetus behind the adoption of HB estimation for choice-based conjoint, our ability to estimate more complex models is limited by the need to customize programs for each new model. For researchers who need to estimate other types of models, WINBUGS and MCMCPack (for the R/S statistical computing environments) offer a means of implementing HB methods.

Practitioners, in particular, must also adjust to some of the differences they will encounter in using HB methods. As one example, HB estimation methods do not "converge" on a closed-form solution in the way that many of our classical estimation methods, such as multinomial logit, do. Practitioners will need to become comfortable with the fact that, once the variance in the estimation stabilizes after several thousand "burn-in" iterations, there will still be considerable variation in the average parameter estimates. Another important difference with HB methods is that, instead of a point estimate of values for each respondent, we usually end up with a distribution of estimates for each respondent. While this is powerful in terms of understanding uncertainty, it adds to the complexity of the analysis, particularly in the case of market simulation.

Applications and Opportunities

HB models are revolutionizing research in marketing. Significant strides have been made in quantifying many aspects of behavior, including sources of measurement error (such as carry-over effects and scale usage), volumetric forecasts, and satiation. Researchers are exploring how individuals differ and how variables capable of reflecting the motivations present when they interact with their environment and seek help through the goods and services available for sale. Finally, HB models have revolutionized the use of spreadsheet simulators that explore marketplace scenarios using the individual-level estimates of utility previously described. While our examples have focused on survey-based data, Bayesian methods have been applied to a wide variety of marketing data, including transaction data and Web site "click through" data.

Bayesian models give researchers in marketing the freedom to study the complexities of human behavior in a more realistic fashion than was previously possible. Human behavior is extremely complex. Unfortunately, many of the models and variables used in our analysis are not. Consider, for example, the linear model in equation 2. While this model has been the workhorse of much statistical analysis, particularly in conjoint analysis, it does not provide a true representation of how respondents encode, judge, and report on items in a questionnaire. The future holds many opportunities to look behind the responses in surveys to gain better insight into how individuals might act in the marketplace. ●

Additional Reading

Gilbride, Timothy J. and Greg M. Allenby (2004), "A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules," *Marketing Science*, forthcoming.

Rossi, Peter E. and Greg M. Allenby (2003), "Bayesian Statistics and Marketing," *Marketing Science*, 22, 304-328.

Greg M. Allenby is the Helen C. Kurtz chair in marketing at the Fisher College of Business, Ohio State University. He may be reached at allenby.1@osu.edu. **David G. Bakken** is senior vice president, marketing science, at Harris Interactive. He may be reached at dbakken@harrisinteractive.com. **Peter E. Rossi** is Joseph T. Lewis professor of marketing and statistics, graduate school of business, University of Chicago. He may be reached at peter.rossi@gsb.uchicago.edu.