

Chapter 4

Introduction to Choice Modeling



4.1 Introduction

Revenue management models were originally developed under the assumption of stochastically independent demands. This assumption is untenable when products are close substitutes. In this case, the demand for a particular product may depend on the set of competing products that are available in the market. For example, when a product is removed from an assortment, its demand may be recaptured by another product in the assortment, or it may spill to competitors or the no-purchase alternative. Conversely, adding a product to the assortment may cannibalize the demand for other products in the assortment or may induce new demand. In this chapter, we study discrete choice models that help capture demand as a function of the offered products. In later chapters, we will use these discrete choice models to formulate optimization problems to choose profit or revenue maximizing assortments when the prices of the products are fixed, or to find the profit or revenue maximizing prices to charge for the offered products. Consequently, when discussing discrete choice models in this chapter, we pay particular attention to those discrete choice models that are rich enough to capture substitution effects and for which the corresponding optimization problems are computationally tractable.

In Sect. 4.2, we describe what we mean by a discrete choice model and formulate assortment and price optimization problems at a high level, but we defer the solution of such problems to later chapters. In Sect. 4.3, we discuss the maximum utility model, which provides a flexible framework to construct various choice models. In Sect. 4.4, we discuss the basic attraction model, of which the well-known multinomial logit model is a special case. In Sect. 4.5, we generalize this model to allow the attraction of the no-purchase alternative to depend on the products excluded from the assortment. A special case of this general attraction model is the independent demand model, where demand for each product is independent of other products offered. In this section, we also discuss the independence of

irrelevant alternatives property, which is a shortcoming of the last two choice models mentioned. In Sect. 4.6, we explain the nested logit model, which alleviates this property, at least to a certain extent. In Sect. 4.7, we demonstrate how we can mix basic attraction models to generate richer choice models. In Sect. 4.8 we present the exponential model based on reflected exponential utilities, while in Sect. 4.9 we study random consideration set models, where the preference order of the products is the same for all consumers, but different consumers drop different products from consideration. Finally, in Sect. 4.10 we discuss a general approach for choice modeling, where each consumer arrives with a particular ordering of products in mind, and she purchases the highest ranking available product. We then present the Markov chain, where consumers purchase their preferred product if available, and otherwise navigate according to a Markov chain until they find an available alternative, which may be the no-purchase alternative. Bounds and approximations are presented in Sect. 4.11 with an interpretation that may fit retail settings better than traditional choice models.

4.2 Discrete Choice Models

We will assume that the set of potential products that could be offered is $N := \{1, \dots, n\}$. For any subset $S \subseteq N$, denote by $\pi_j(S)$ the probability that a consumer will select product $j \in S$, with $\pi_j(S) = 0$ if $j \notin S$. Let $\Pi(S) := \sum_{j \in S} \pi_j(S)$ denote the probability of a sale when subset S is offered. The complement $\pi_0(S) := 1 - \Pi(S)$ denotes the probability that the consumer selects an outside alternative. An outside alternative may mean either that the consumer does not purchase or that she purchases from another vendor. The outside alternative is always implicitly available. For this reason, it would be more appropriate to write $\pi_j(S_+)$ for $j \in S_+ := S \cup \{0\}$. However, we follow here the convention of writing $\pi_j(S)$ instead of the more cumbersome $\pi_j(S_+)$. Notice that under a discrete choice model a consumer will select exactly one product from the set S_+ when assortment S is offered. This is appropriate for transportation and lodging choices. Later we describe models that relax this assumption that may be more appropriate in a retail setting where consumers may buy more than one product.

A tedious way to describe any choice model is to list $\pi_j(S)$ for each $j \in S$ and for each $S \subseteq N$. This approach would require specifying $n 2^{n-1}$ scalars which are far too many parameters to list or estimate. This is primarily why researchers have focused on parsimonious choice models that are based on a few parameters typically of the order $O(n)$ or $O(n^2)$. Most of the parametric choice models described in this chapter are parsimonious in nature.

In addition to our interest in finding realistic choice models to describe demand, we are also interested in finding a subset of products, say $S \subseteq N$, to offer to consumers with the objective of maximizing expected profits or revenues. Let p_j and z_j be, respectively, the price and the unit cost of product $j \in N$, and define the vectors $p := (p_1, \dots, p_n)$ and $z := (z_1, \dots, z_n)$. Then the expected profit

obtained by offering the subset S is given by $R(S, z) := \sum_{j \in S} (p_j - z_j) \pi_j(S)$. The assortment optimization problem is that of finding a subset $S \subseteq N$ that maximizes $R(S, z)$, yielding

$$\mathcal{R}(z) := \max_{S \subseteq N} R(S, z). \quad (4.1)$$

The assortment optimization problem is combinatorial in nature and arises frequently in retailing and revenue management. We take on the question of formulating and solving such problems in the next chapter on assortment optimization.

If the prices are also decision variables, then the choice model needs to be price aware. Let $\pi_i(N, p)$ be the probability of selecting product i when the set N is offered. We are interested in maximizing $R(p, z) := \sum_{j \in N} (p_j - z_j) \pi_j(N, p)$ over all vectors $p \geq z$, yielding

$$\mathcal{R}(z) := \max_{p \geq z} R(p, z).$$

For many choice models, if $p_j = \infty$, then $\pi_j(N, p) = 0$. Thus, setting $p_j = \infty$ is equivalent to not offering product j , indicating that optimizing over prices could implicitly select an assortment of products to offer. Problems of the form above are nonlinear optimization problems that arise in dynamic pricing and will also be encountered in a later chapter.

4.3 Maximum and Random Utility Models

Suppose there is a full preference ordering among the products. We assume without loss of generality that the products are labeled so the preferences are $1 \prec 2 \prec \dots \prec n$. The maximum utility model (MUM) assigns $\pi_i(S) = 1$ for $i \in S$, if and only if $j \in S, j \neq i$ implies $j \prec i$. In other words, consumers select the highest ranked product in S , which corresponds to the product with the highest utility in the set S . The MUM is often presented by assigning cardinal utilities, say $\{u_i : i \in N\}$ to the products. If the u_i 's are all different, then we can order them $u_1 < u_2 < \dots < u_n$, and for any $S \subseteq N$ the maximum utility $\max_{i \in S} u_i$ is attained by the highest ranked product in S . If instead $u_1 \leq u_2 \leq \dots \leq u_n$, then there may be more than one product in S attaining $\max_{i \in S} u_i$, in which case choice probabilities are assigned uniformly among such products.

Random utility models (RUM) add a random noise component to the utilities of the products, $U_i = u_i + \epsilon_i, i \in N$, where the ϵ_i 's are mean zero, possibly dependent random variables. Another way to introduce randomness into the MUM is to assume a distribution of consumer types, each with a certain preference ordering. So, product $i \in S$ is selected by types of consumers for whom i is the highest ranked product in S .

All RUM's satisfy the regularity property that the probability of choosing a product does not increase as other products are added to the assortment. There are known cases where the regularity property does not hold, and readers are cautioned that in such cases there is a cost of approximating the choice model by a RUM. Nevertheless, RUM's are an important class to which we devote most of the chapter. We provide references to more general discrete choice models at the end of the chapter.

4.4 Basic Attraction and Multinomial Logit Models

The basic attraction model (BAM) is a discrete choice model where each product $j \in N$ has an attraction value $v_j > 0$, capturing the attractiveness of product j to a consumer. Similarly, the attraction value $v_0 > 0$ represents the attractiveness of the no-purchase alternative. The choice model is given by

$$\pi_j(S) = \frac{v_j}{v_0 + V(S)} \quad \forall j \in S, \quad (4.2)$$

where $V(S) := \sum_{j \in S} v_j$. Consequently, products with higher attraction values are more likely to be selected.

The introduction of BAM is based on postulating two choice axioms and demonstrating that a discrete choice model satisfies the axioms if and only if it is of the BAM form. To describe these axioms, we need additional notation. For any $S \subseteq T$, let $\pi_S(T) := \sum_{j \in S} \pi_j(T)$ denote the probability that a consumer selects a product in S when the set T is offered. Also, $\pi_{S_+}(T) := \pi_S(T) + \pi_0(T) = 1 - \pi_{T \setminus S}(T)$, where $T \setminus S$ is the set difference of T from S . The Luce axioms can be written as follows:

- Axiom 1: If $\pi_i(\{i\}) \in (0, 1)$ for all $i \in T$, then for any $Q \subseteq S_+$, $S \subseteq T$

$$\pi_Q(T) = \pi_Q(S) \pi_{S_+}(T).$$

- Axiom 2: If $\pi_i(\{i\}) = 0$ for some $i \in T$, then for any $S \subseteq T$ such that $i \in S$

$$\pi_S(T) = \pi_{S \setminus \{i\}}(T \setminus \{i\}).$$

Axiom 1 implies that the probability of selecting any set $Q \subseteq S_+$, when set T is offered, is equal to the probability of selecting Q when S is offered times the probability of selecting S_+ when T is offered assuming that $S \subseteq T$. Axiom 2 implies that if alternative i has no probability of being chosen, then it can be deleted without affecting the choice probabilities.

The celebrated multinomial logit (MNL) model is a special case of the BAM that arises from a RUM. Under a RUM, each product j has a random utility $U_j =$

$u_j + \epsilon_j$ for $j \in N_+$ and the probability that product $j \in S$ is selected is given by $\pi_j(S) = \mathbb{P}\{U_j \geq U_i \forall i \in S_+\}$. We can think of ϵ_j as an idiosyncratic variation on the mean utility or as errors in measuring the utility.

If ϵ_j is a normal random variable, then the resulting choice model is known as the Probit model. Unfortunately, there is no closed-form expression for the selection probabilities under the Probit model and that has discouraged its use to a certain extent. However, a closed-form expression for the selection probabilities can be obtained if $\{\epsilon_j : j \in N_+\}$ are modeled as independent and identically distributed Gumbel random variables all having the same scale parameter. The cumulative distribution function of a Gumbel random variable X with location and scale parameters ν and ϕ is given by

$$F(x : \nu, \phi) = \exp(-\exp(-\phi(x - \nu))). \quad (4.3)$$

The mode of this distribution is ν , while the median is $\nu - \ln(\ln 2)/\phi$. The mean is $\mathbb{E}[X] = \nu + \gamma/\phi$, where γ is the Euler constant. The variance is given by $\text{Var}[X] = \pi^2/6\phi^2$, where π is the ratio of a circle's circumference to its diameter. To obtain a mean zero random variable, we set $\nu = -\gamma/\phi$. Notice that the variance is inversely proportional to ϕ^2 . If $\{\epsilon_j : j \in N_+\}$ are mean zero Gumbel random variables all with scale parameter ϕ and independent across the products, then

$$\pi_j(S) = \frac{e^{\phi u_j}}{1 + \sum_{k \in S} e^{\phi u_k}} \quad \forall j \in S,$$

and this is known as the MNL model, and a special case of the BAM.

As ϕ becomes large, the variance of ϵ_j becomes small and the choice probabilities concentrate on the product or products with the largest mean utility. Thus, only the products with the largest mean utilities are purchased resulting in the MUM. On the other hand, when ϕ becomes small, the probability of selecting any offered product converges to a uniform distribution, where each product is equally likely to be selected. This behavior arises because when the variance is much larger than the mean, the consumer loses the ability to reliably select products with higher mean utility.

4.5 Generalized Attraction Model

One of the shortcomings of the BAM is that the attraction value v_0 of the no-purchase option is a fixed parameter and does not depend on the subset of offered products. There is considerable empirical evidence that the BAM may be too optimistic in estimating demand recapture probabilities when the first choice of a consumer is not part of the offer set S . In particular, the BAM assumes that even if a consumer prefers some product $j \in \bar{S} := N \setminus S$, she must select a product in S_+ . This approach ignores the possibility that the consumer may look for the products in \bar{S} elsewhere or at a later time.

As an example, suppose that a consumer prefers a certain wine, and the store does not have it. The consumer may then either buy one of the wines in the store, go home without purchasing, or drive to another store and look for the specific wine she wants. The BAM precludes the last possibility; it implicitly assumes that the search cost for an alternative source of product $j \in \bar{S}$ is infinity, or equivalently that there is no competition. As an illustration, suppose that the consideration set is $N = \{1, 2\}$ and that $v_0 = v_1 = v_2 = 1$, so $\pi_j(\{1, 2\}) = 33.3\%$ for $j = 0, 1, 2$. Under the BAM, eliminating choice 2 results in $\pi_j(\{1\}) = 50\%$ for $j = 0, 1$. Suppose, however, that product 2 is available across town and the attraction value for product 2 from the alternative source is $w_2 = 0.5$. Then the choice set of a consumer, when product 2 is not offered, is in reality $\{1, 2'\}$ with $2'$ representing product 2 in the alternative location with shadow attraction w_2 . Under this model,

$$\pi_0(\{1\}) = \frac{1.5}{2.5} = 60\%, \quad \pi_1(\{1\}) = \frac{1}{2.5} = 40\%.$$

To formally define the generalized attraction model (GAM), we assume that in addition to the attraction values $\{v_j : j \in N\}$, there are shadow attraction values $\{w_j : j \in N\}$ with $w_j \in [0, v_j]$ for all $j \in N$. Under the GAM, for any subset $S \subseteq N$, the selection probabilities are given by

$$\pi_j(S) = \frac{v_j}{v_0 + W(\bar{S}) + V(S)} \quad \forall j \in S, \quad (4.4)$$

where $W(R) := \sum_{j \in R} w_j$ for any $R \subseteq N$. Thus, the GAM can be viewed as a version of the BAM, where the attractiveness of the no-purchase alternative is inflated to $v_0 + \sum_{j \in \bar{S}} w_j$ as a function of the shadow attraction values of the alternatives that are not offered. The no-purchase probability is given by

$$\pi_0(S) = \frac{v_0 + W(\bar{S})}{v_0 + W(\bar{S}) + V(S)}.$$

The GAM can be obtained axiomatically by modifying one of the Luce axioms. The special case of the GAM with $w_j = 0$ for all $j \in N$ recovers the BAM. As with the BAM, it is possible to normalize the parameters so that $v_0 = 1$ when $v_0 > 0$.

The parsimonious GAM (p-GAM) is given by $w_j = \theta v_j$ for all $j \in N$ for some $\theta \in [0, 1]$. In this model, the shadow attraction values are a constant factor of the original attraction values. The special case $\theta = 0$ recovers the BAM.

At the other extreme, the case $\theta = 1$ results in the independent demand model (IDM). Under the IDM, the probability $\pi_j(S)$ of selecting product $j \in S$ is independent of the offer set S , as long as S includes product j . If product j is removed, then all of its demand is lost to the no-purchase alternative. This implies that there are nonnegative constants, say v_0 and $\{v_j : j \in N\}$, such that

$$\pi_j(S) = v_j \quad \forall j \in S,$$

and $v_0 + V(N) = 1$. The IDM model may reflect what happens in very competitive situations where consumers can readily find another vendor offering products not in S . In most practical situations, however, it is reasonable to expect that some of the demand for product $j \notin S$ may be recaptured by other products in S .

The p-GAM can serve to test the competitiveness of the market, by testing the hypothesis $H_0 : \theta = 0$ or $H_0 : \theta = 1$, to see whether the BAM applies or the independent demand model applies. To test these hypotheses, we can use offer sets S_t and realized sales s_t over a certain period of time $t = 1, \dots, T$ and obtain the likelihood functions $L(v, \theta)$. We would reject $H_0 : \theta = 0$ in favor of $H_1 : \theta = \theta_1 > 0$ if $\max_v L(v, 0) / \max_v L(v, \theta_1)$ is sufficiently small. Likewise, we would reject $H_0 : \theta = 1$ against $H_1 : \theta = \theta_1 < 1$ if $\max_v L(v, 1) / \max_v L(v, \theta_1)$ is sufficiently small. If these tests fail, then a GAM maybe a better fit to the data.

There is an alternative, perhaps simpler, way of presenting the GAM by using the following transformation. For all $j \in N$, we let $\tilde{v}_j = v_j - w_j$ and $\tilde{v}_0 = v_0 + W(N)$. For $S \subseteq N$, let $\tilde{V}(S) = \sum_{j \in S} \tilde{v}_j$. With this notation, the selection probabilities under the GAM given in (4.4) can equivalently be written as

$$\pi_j(S) = \frac{v_j}{\tilde{v}_0 + \tilde{V}(S)} \quad \forall j \in S \quad \text{and} \quad \pi_0(S) = 1 - \sum_{j \in S} \pi_j(S).$$

The form of the selection probability above is similar to the one under the BAM given in (4.2). This similarity becomes useful when extending assortment optimization results for the BAM to the GAM.

4.5.1 Independence of Irrelevant Alternatives

A close inspection of the selection probabilities under the BAM and GAM reveals that if we add a new product to an offered subset, then the purchase probability of all offered products decreases by the same relative amount. In particular, for any set $S \subseteq N$ and product $j, i \in S$ and $k \in \bar{S}$, the selection probabilities under the BAM and GAM satisfy

$$\frac{\pi_j(S)}{\pi_j(S \cup \{k\})} = \frac{\pi_i(S)}{\pi_i(S \cup \{k\})}.$$

The expression on the left can be interpreted as the relative decrease in the purchase probability of product j when we add product k to the offered set, whereas the expression on the right can be interpreted as the relative decrease in the purchase probability of product i when we add product k to the offered set. This situation is referred to as the independence of irrelevant alternatives (IIA) property. It can equivalently be stated as the ratio $\pi_i(S) / \pi_j(S)$ is independent of the set S containing both products i and j . For the BAM and the GAM, this ratio is v_i / v_j .

Intuitively, the IIA property should not hold when k cannibalizes the demand for products j and i to different extents. To see the negative consequences that the

IIA can have on the BAM, we use the well-known red bus, blue bus paradox. In this paradox, a person has a choice between driving a car and taking either a red or a blue bus. It is implicitly assumed that both buses have ample capacity and depart at the same time. Let v_c represent the attraction value of driving a car and let v_b represent the attraction value of riding the red bus. Under the BAM with $v_0 = 0$, the probability of driving, when the choice set is between driving and riding the red bus, is $v_c/(v_c + v_b)$. Adding a blue bus with attraction value v'_b drops the probability of driving to $v_c/(v_c + v_b + v'_b)$. This drop has been widely criticized because the addition of an equally attractive blue bus in addition to the red bus should not influence the probability of driving to the extent it does under the BAM.

While the GAM also suffers from the IIA property, it can better handle the blue bus, red bus paradox. Indeed, by setting $w_b = v'_b$, the probability of driving remains unchanged by the introduction of the blue bus under the GAM as long as $v'_b \leq v_b$. More generally, the probability of driving can be modeled as $v_c/(v_c + \max(v_b, v'_b))$ as driving competes with the more attractive of the two buses. We can fit this into the GAM framework by setting $v_b \leftarrow \max(v_b, v'_b)$ and $w_b \leftarrow \min(v_b, v'_b)$. A more common fix for the IIA property is to use the nested logit model, which we describe next.

4.6 Nested Logit Model

In the nested logit (NL) model, the products are organized into nests such that the products in the same nest are regarded as closer substitutes of each other relative to the products in different nests. Under the NL model, the selection process of a consumer proceeds in two stages. First, the consumer selects either one of the nests or decides to leave without making a purchase. Second, if the consumer selects one of the nests, then the consumer chooses one of the products offered in this nest. To formulate the NL model, we use $M := \{1, \dots, m\}$ to denote the set of nests. For notational brevity, we assume that the number of products in each nest is the same and we use N to denote the set of products available in each nest. It is straightforward to generalize our formulation to the case where different nests have different numbers of products. We use $S_i \subseteq N$ to denote the set of products offered in nest i . Therefore, the sets of products offered over all nests are given by $\{S_1, \dots, S_m\}$. The attraction value of product j in nest i is given by v_{ij} . Under the NL model, if a consumer has already decided to make a purchase in nest i and the set $S_i \subseteq N$ of products is offered in this nest, then the consumer selects product $j \in S_i$ with probability

$$q_{j|i}(S_i) := \frac{v_{ij}}{V_i(S_i)},$$

where $V_i(S_i) := \sum_{j \in S_i} v_{ij}$. On the other hand, each nest i has a dissimilarity parameter $\gamma_i \in [0, 1]$. The parameter γ_i is a measure of how easily the products in nest i substitute for each other. In this case, if the sets of products offered over all

nests are given by $\{S_1, \dots, S_m\}$, then a consumer chooses nest i with probability

$$Q_i(S_1, \dots, S_m) := \frac{V_i(S_i)^{\gamma_i}}{v_0 + \sum_{l \in M} V_l(S_l)^{\gamma_l}},$$

where v_0 denotes the attraction value of the no-purchase alternative. Thus, if we offer the sets of products $\{S_1, \dots, S_m\}$, then the selection probability of product j in nest i is given by

$$Q_i(S_1, \dots, S_m) q_{j|i}(S_i) = \frac{V_i(S_i)^{\gamma_i}}{v_0 + \sum_{l \in M} V_l(S_l)^{\gamma_l}} \frac{v_{ij}}{V_i(S_i)}.$$

In Sect. 4.4, we discuss that the MNL model can be interpreted as a RUM, where a consumer associates random utilities with the options and chooses the option providing the largest utility. The NL model has the same kind of a random utility interpretation. In particular, assume that a consumer associates the utility $U_{ij} = u_{ij} + \epsilon_{ij}$ with product j in nest i , where u_{ij} and ϵ_{ij} are respectively the deterministic and random utility components. We assume that $\epsilon = \{\epsilon_{ij} : i \in M, j \in N\}$ has a type of generalized extreme value distribution and the joint distribution function for ϵ is given by

$$F(x; \gamma) = \exp \left(- \sum_{i \in M} \left(\sum_{j \in N} \exp(-x_{ij}/\gamma_i) \right)^{\gamma_i} \right),$$

where we use x and γ to denote the vectors $\{x_{ij} : i \in M, j \in N\}$ and $\{\gamma_i : i \in M\}$. With the generalized extreme value distribution above, the marginal distribution of ϵ_{ij} is Gumbel and has the form (4.3). For two distinct nests $l, k \in M$, the random utilities ϵ_{ij} and ϵ_{lk} are independent of each other, but for a given nest i , the random utilities ϵ_{ij} and ϵ_{ik} are positively correlated. The parameter $1 - \gamma_i$ measures the degree of correlation between the utilities in nest i . If the random utilities have this form of correlated generalized extreme value distribution and consumers choose the product that provides the largest utility, then the selection probabilities under the corresponding RUM have the same form as the selection probabilities under the nested logit model, when the attraction values are $v_{ij} = e^{u_{ij}/\gamma_i}$ for all $i \in M, j \in N$.

The RUM interpretation of the NL model explains why products in the same nest are closer substitutes of each other. If two products are in the same nest, then their utilities are positively correlated. Thus, if a consumer associates a high utility with product j in nest i , then this consumer is also likely to associate a high utility with product k in nest i . In this case, if product j is not available but product k is available, then the consumer is likely to substitute product k for product j . As γ_i approaches to zero, the utilities of the products in the same nest become more positively correlated, so they become closer substitutes. When $\gamma_i = 1$ for all $i \in M$, the utilities of the products are uncorrelated and the NL model reduces to the MNL model. In some settings, the NL model has been used with dissimilarity parameters

$\{\gamma_i : i \in M\}$ exceeding one. This form of the NL model does not necessarily have a random utility interpretation. However, when one estimates the parameters of the NL model from the data, it is conceivable to end up with estimates of the dissimilarity parameters exceeding one that perform well both with training and testing data. Consequently, models with $\gamma_i > 1$ are often fit in practice even if they lose their RUM interpretation.

One of the attractive features of the BAM, GAM, and NL model is that the assortment optimization problem formulated in (4.1) is tractable when consumers choose according to one of these choice models. In the later chapters, we show how to solve problem (4.1) under these choice models.

4.7 Mixtures of Basic Attraction Models

One option to add richness to the BAM is to consider a version of BAM with multiple consumer segments. This adds heterogeneity in tastes. In particular, we assume that there are multiple consumer types and we use G to denote the set of all possible consumer types. Customers of type g associate the attraction value v_j^g with product j and the attraction value v_0^g with the option of not making a purchase. An arriving consumer is of type g with probability α^g . In this case, if we offer the set S of products, then the selection probability for product $j \in S$ is

$$\pi_j(S) = \sum_{g \in G} \alpha^g \frac{v_j^g}{v_0^g + \sum_{i \in S} v_i^g} \quad \forall j \in S.$$

The choice model above is referred to as a mixture of BAMs. It can be shown that any discrete choice model that arises from a RUM can be approximated to any degree of accuracy by a mixture of BAM's. This result indicates that the mixture of BAM's can be quite flexible in modeling a variety of consumer choice patterns. However, as we will see in the next chapter, the mixture of BAMs leads to difficulties in assortment optimization when we try to find a revenue maximizing set of products to offer to consumers. In Sect. 4.10 we describe an alternative choice model that can provide a good approximation to the mixture of BAM's, while keeping the corresponding assortment optimization problem tractable.

4.8 The Exponential Model

In the exponential model, the utility of alternative i is $U_i = u_i + \epsilon_i$ for $i \in N_+$, where the ϵ_i 's are independent, mean zero random variables of the form $\epsilon_i = \theta - \tau_i$ and τ_i is an exponential random variable with mean θ . Suppose that a subset $S \subseteq N$ is offered, the cardinality of S_+ is m , and that the products in S_+ are sorted in increasing order of u_i . Let $\{1, \dots, m\}$ be the label of the products in S_+ under this

sorting. Let $G(0) = 0$, and

$$G(i) := \frac{\exp(-\sum_{j=i}^m (u_j - u_i)/\theta)}{m - i + 1} \quad \text{for } i \in \{1, \dots, m\}.$$

Then, the choice probabilities are given as follows:

$$\pi_i(S_+) = G(i) - \sum_{j=1}^{i-1} \frac{G(j)}{m - i} \quad i \in \{1, \dots, m\},$$

where sums over empty sets are assumed to be zero. Notice that in this model, the no-purchase alternative is one of the elements in $\{1, \dots, m\}$. On the surface, the exponential model seems difficult to work with, but on the positive side the parameters are easy to estimate.

4.9 Random Consideration Set Model

The random consideration set (RCS) model is characterized by a strict preference order $<$ on N , and by a vector λ of positive attention probabilities. We assume without loss of generality that the products are labeled so that $1 < 2 < \dots < n$. Then for any subset $S \subseteq N$, the choice probability of product i is

$$\pi_i(S) = \lambda_i \prod_{j>i, j \in S} (1 - \lambda_j) \quad \forall i \in S, \quad (4.5)$$

with $\pi_i(S) = 0$ if $i \notin S$. The interpretation of the choice probability above is that a consumer forms a consideration set $C(S)$ by independently including each product i in her consideration set with probability λ_i . She then selects the most preferable available product in her consideration set. Thus, for a consumer to purchase product i , product i should be in her consideration set and all offered products that are preferred to product i should not be in her consideration set. In other words, $i \in S$ is chosen if and only if i is the highest ranked product in the random consideration set $C(S)$. The extreme case $\lambda_j = 1$ for all $j \in N$ corresponds to the MUM under a strict preference order. This case corresponds to having $\pi_i(S) = 1$ if and only if $i > j$ for all $j \in S$.

The RCS model assumes implicitly that $\lambda_0 = 1$ and that consumers pay no attention to products that are ranked below the no-purchase alternative, so consumers select the no-purchase alternative among assortments $S \subseteq N$ only from inattention to the alternatives in the assortment. Indeed,

$$\pi_0(S) = \lambda_0 \prod_{j \in S} (1 - \lambda_j) = \prod_{j \in S} (1 - \lambda_j) > 0,$$

if and only if all the products $j \in S$ have inattention probabilities $1 - \lambda_j > 0$.

Product cannibalization is asymmetric in this model, as the introduction of product i to an assortment S cannibalizes the demand of products $j \in S$ only if $j < i$. This can be seen by noting that $\pi_j(S \cup \{i\}) = \pi_j(S)$ if $i < j$, while $\pi_j(S \cup \{i\}) = \pi_j(S)(1 - \lambda_i)$ if $i > j$. In words, demand for a product can only be cannibalized by a product higher up in the preference ordering. The cannibalization asymmetry exhibited by the RCS model is an important feature that is difficult to capture by other choice models. Thus, the RCS model is particularly useful when there is evidence of cannibalization asymmetry. It turns out that the RCS model is a special case of the Markov chain choice model presented next.

4.10 Markov Chain Choice Model

A random utility model can be viewed as a probability distribution over all preference lists, or permutations, over $N_+ = N \cup \{0\}$. A permutation σ is a bijection from $N_+ \rightarrow N_+$ resulting in the preference ordering $\sigma(1) > \sigma(2) > \dots > \sigma(n+1)$. Each permutation σ of N_+ has a certain probability, say $\mathbb{P}(\sigma) \geq 0$, with $\sum_{\sigma} \mathbb{P}(\sigma) = 1$. The probability that consumers prefer product i when the full set of options N is offered is given by

$$\lambda_i := \pi_i(N) = \sum_{\sigma} \mathbb{P}\{\sigma(1) = i\}, \quad \forall i \in N_+, \quad (4.6)$$

where $\sigma(\ell)$ is the product with rank ℓ in the permutation σ . In words, the choice probability $\pi_i(N)$ is obtained by adding over all of those permutations that have product i in the first position. We refer to the λ_i as the first choice probability of product $i \in N$, as it is the probability of selecting product i when the entire set of products N is offered. Clearly, $\sum_{i \in N_+} \lambda_i = 1$.

If product i is not available, then consumers whose first choice is i substitute to product $j \neq i$ with probability

$$\rho_{ij} = \sum_{\sigma} \mathbb{P}\{\sigma(2) = j \mid \sigma(1) = i\} \quad \forall i \neq j, i \in N, j \in N_+.$$

In words, ρ_{ij} is the conditional probability that a consumer whose first choice is $i \in N$ will have $j \in N_+$ as its second choice. Because the no-purchase alternative is always available, we set $\rho_{0,j} = 0$ for all $j \in N$, and $\rho_{0,0} = 1$. Notice that $\sum_{j \in N_+} \rho_{ij} = 1$ for all $i \in N_+$.

Assuming that we know $\lambda_i = \pi_i(N)$ and $\pi_i(N \setminus \{j\})$ for all $i, j \in N$, we can compute the substitution probabilities using the formula

$$\rho_{ij} = \frac{\pi_j(N \setminus \{i\}) - \pi_j(N)}{\pi_i(N)} \quad \forall i \neq j, i \in N, j \in N_+.$$

If we knew the distribution $\mathbb{P}(\sigma)$ over all permutations, we could compute $\pi_i(S)$ by modifying (4.6) so that the sum includes all permutations where i appears before any other element in $S_+ = S \cup \{0\}$. This is a lot of work, even if we knew $\mathbb{P}(\sigma)$ for all σ . Suppose instead, that we try to approximate $\pi_i(S)$ based only on the information contained in $\lambda = \{\lambda_i : i \in N\}$ and $\rho = \{\rho_{ij} : i, j \in N\}$. To do this we need to make an assumption about what happens if a consumer does not find her first choice, say $i \notin S$, and transitions to some $j \in N_+$. If the transition is to 0, then the consumer will leave the system without purchase. If the transition is to $j \in S$, then the consumer will purchase product j as this is her second choice after i . When $j \in \bar{S}$ we make the Markovian assumption that the consumer will continue evolving according to a Markov chain as if her first choice was product j . She would then transition to product k with probability ρ_{jk} until she either transitions to a product in S or to the no-purchase alternative. This choice process corresponds to the one under the Markov chain (MC) choice model. Thus, the MC choice model can be interpreted as a permutation-based choice model, but once a consumer considers a particular product in her preference list, she loses track of the earlier products in the preference list.

Example 4.1 If $n = 2$, then there are six permutations of $N_+ = \{0, 1, 2\}$. If the probabilities of observing these permutations are given by $\mathbb{P}(0, 1, 2) = 0.2$, $\mathbb{P}(0, 2, 1) = 0.15$, $\mathbb{P}(1, 2, 0) = 0.2$, $\mathbb{P}(1, 0, 2) = 0.1$, $\mathbb{P}(2, 1, 0) = 0.15$, and $\mathbb{P}(2, 0, 1) = 0.2$, then $\lambda_1 = \mathbb{P}(1, 2, 0) + \mathbb{P}(1, 0, 2) = 0.3$, $\lambda_2 = \mathbb{P}(2, 1, 0) + \mathbb{P}(2, 0, 1) = 0.35$, and $\lambda_0 = \mathbb{P}(0, 1, 2) + \mathbb{P}(0, 2, 1) = 0.35$. Moreover, we have $\rho_{1,2} = \mathbb{P}\{\sigma(2) = 2 | \sigma(1) = 1\} = \mathbb{P}(\sigma(1, 2, 0)) / \mathbb{P}\{\sigma(1) = 1\} = 2/3$ and consequently $\rho_{1,0} = 1/3$. Similarly, $\rho_{2,1} = \mathbb{P}\{\sigma(2) = 1 | \sigma(1) = 2\} = \mathbb{P}(\sigma(2, 1, 0)) / \mathbb{P}\{\sigma(1) = 2\} = 0.15/0.3 = 1/2$, so $\rho_{2,0} = 1/2$. The full specification of the MC choice model is given by $\lambda = (0.3, 0.35, 0.35)$ and

$$\rho = \begin{pmatrix} 1 & 0 & 0 \\ 1/3 & 0 & 2/3 \\ 1/2 & 1/2 & 0 \end{pmatrix}.$$

Clearly $\pi_2(\{2\}) = \mathbb{P}(1, 2, 0) + \mathbb{P}(2, 1, 0) + \mathbb{P}(2, 0, 1) = 0.2 + 0.15 + 0.2 = 0.55$ as these are the three permutations where 2 is preferred to 0. However, this calculation requires the knowledge of the distribution among permutations. We can estimate $\pi_2(\{2\})$ from λ and ρ as follows. With probability $\lambda_1 = 0.3$ consumers prefer product 1 and among those $2/3$ will transition to product 2. In addition, there is a $\lambda_2 = 0.35$ probability that consumers directly prefer product 2. Then, our estimate of $\pi_2(\{2\})$ is given by $\hat{\pi}_2(\{2\}) = 0.3 \times (2/3) + 0.35 = 0.55$ which is exactly $\pi_2(\{2\})$. Similarly, we can compute $\pi_1(\{1\}) = \mathbb{P}(1, 2, 0) + \mathbb{P}(1, 0, 2) + \mathbb{P}(2, 1, 0) = 0.2 + 0.1 + 0.15 = 0.45$, while the MC approximation results in $\hat{\pi}_1(\{1\}) = 0.3 + 0.35 \times (1/2) = 0.475$.

It is possible to write a system of equations to describe the selection probabilities under the MC choice model. We will denote by $\phi_j(S)$ the probability that a

consumer considers product $j \notin S$ during the course of her choice process but does not purchase because $j \notin S$. By definition, $\phi_j(S) = 0$ for all $j \in S$. The quantities $\pi_j(S)$ for $j \in S$ and $\phi_j(S)$ for $j \in \bar{S}$ are related by the system of equations

$$\pi_j(S) = \lambda_j + \sum_{i \in \bar{S}} \phi_i(S) \rho_{ij} \quad \forall j \in S \quad (4.7)$$

$$\phi_j(S) = \lambda_j + \sum_{i \in \bar{S}} \phi_i(S) \rho_{ij} \quad \forall j \in \bar{S}. \quad (4.8)$$

The system of equations in (4.7) and (4.8) above can be interpreted as balance equations and are simply a reinterpretation of the steady-state probabilities. Considering the first set of equations, a consumer ends up purchasing product $j \in S$ either as a first choice with probability λ_j , or by visiting some product $i \in \bar{S}$ during the course of her choice process, not purchasing this product because it is not available, and by transitioning to product j . Similarly, for $j \notin S$, a consumer may consider product j as she enters the system, with probability λ_j , or after transitioning to j from other product $i \in \bar{S}$. The sets of equations in (4.7) and (4.8) consist of $|N|$ equations, which we can solve to obtain the $|N|$ probabilities $\{\pi_j(S) : j \in S\}$ and $\{\phi_j(S) : j \in \bar{S}\}$, with $\pi_0(S) = 1 - \sum_{j \in S} \pi_j(S)$.

Consider the BAM with parameters v_0, v_1, \dots, v_n , normalized so that $\sum_{i \in N_+} v_i = 1$, then

$$\lambda_i = v_i, \quad \forall i \in N_+, \quad \text{and} \quad \rho_{ij} = v_j / (1 - v_i) \quad \forall j \neq i, i \in N, j \in N_+. \quad (4.9)$$

In this case, the solution of the system of equations in (4.7)–(4.8) is $\pi_i(S) = v_i / (v_0 + V(S))$ for all $i \in S$ and $\phi_i(S) = v_i(1 - v_i) / (v_0 + V(S))$ for all $i \notin S$. This solution for $\pi_i(S)$ is exactly what the BAM would predict. Consequently, in the case of the BAM, all of the information about the choice model is contained in (λ, ρ) . It can be shown that the GAM is also a special case of a MC choice model with a rank-one transition matrix, and the RCS model is a special case of the MC choice model with a rank-one triangular transition matrix. In a later chapter we will argue that the linear demand model $d(p) = a - Bp$, with $d(p) \geq 0$ can also be viewed as MC choice model when only products in S are allowed.

The MC choice model can also be used to approximate the mixture of BAM's, where consumers of type $g \in G$ choose according to a BAM with nonnegative attraction values $\{v_j^g : j \in N_+\}$ normalized so $v_0^g + \sum_{j \in N} v_j^g = 1$ for all $g \in G$. Consequently, the first choice and substitution probabilities λ_i^g and ρ_{ij}^g for type g consumers are of the form (4.9). Assume that a consumer is of type g with probability $\alpha^g > 0$, where we have $\sum_{g \in G} \alpha^g = 1$. The formulas for the first choice and substitution probabilities λ_i and ρ_{ij} for the mixed model are given by

$$\lambda_i = \sum_{g \in G} \alpha^g v_i^g, \quad i \in N_+ \quad \text{and} \quad \rho_{ij} = \sum_{g \in G} \alpha^{g|i} \rho_{ij}^g, \quad j \neq i, i \in N, j \in N_+,$$

where

$$\alpha^{g|i} := \frac{\alpha^g v_i^g}{\sum_{c \in G} \alpha^c v_i^c} \quad \forall g \in G, \forall i \in N$$

is the conditional probability that a consumer whose first choice is i is of type g .

On the surface, the MC choice model seems difficult to work with because even computing the selection probabilities requires solving a system of equations. In the next chapter, we will demonstrate that the revenue-maximizing set of products to offer to consumers under this choice model can actually be computed in an efficient manner. Moreover, the MC choice model can be used to approximate any discrete choice model. For example, the model can be fitted by obtaining unbiased guesstimates of the transition probabilities $\{\rho_{ij} : i, j \in N\}$ from a group of managers, with those estimates perhaps averaged with weights that reflect experience in doing such estimations. Alternatively, if there is a metric for the distance between products, then the transition probabilities can be fitted as a decaying function of that metric. More precisely, if δ_{ij} is the distance between products i and j , then ρ_{ij} can be modeled as $\rho_{ij} = e^{-\beta \delta_{ij}}$ for some $\beta > 0$, which can be calibrated by making sure that $1 - \sum_{j \in N \setminus \{i\}} e^{-\beta \delta_{ij}}$ matches the probability of losing the consumer to the no-purchase alternative. The flexibility of the MC choice model together with the ease with which it is possible to find a revenue- or profit-maximizing assortment makes it a useful tool both as a choice model and as a mechanism to find optimal or near-optimal assortments.

Empirical evidence suggests that different choice models can be effective in fitting data. A parsimonious model such as the BAM can be too inflexible to capture choice behavior accurately, while a mixture of BAM's may be too flexible and may suffer from large errors on test data. The MC choice model, particularly the rank-one versions like the GAM, and triangular versions such as the RCS model have a good mixture of flexibility while still being fairly parsimonious.

4.11 Bounds and Approximate Choice Probabilities

Consider a random utility model where $U_i = u_i + \epsilon_i$ for all $i \in N_+$, $\mathbb{E}[\epsilon_i] = 0$ for all $i \in N_+$, and $\epsilon = (\epsilon_0, \epsilon_1, \dots, \epsilon_n)$ follows an absolutely continuous joint distribution that is independent of u_i 's. For any $S \subseteq N$, let

$$G(u, S) := \mathbb{E}[\max_{i \in S_+} U_i]$$

be the expected surplus for the consumer when the offered assortment is S . The Williams-Daly-Zachary theorem

$$\pi_i(S) = \frac{\partial G(u, S)}{\partial u_i} = \mathbb{P}(U_i \geq U_j, \forall j \neq i, j \in S_+)$$

gives the choice probabilities for all $i \in S_+$.

With the exception of a few models like the ones discussed in this chapter, computing $G(u, S)$ and $\pi_i(S)$ $i \in S_+$ can be quite difficult. In what follows, we obtain a bound on $G(u, S)$ and an approximation to $\pi_i(S)$ that is easy to compute even if the ϵ_i 's are not independent. These approximations are fairly accurate and quite useful for cases where there are no closed-form solutions. The approximations themselves can be considered bona fide choice models as the examples below will show.

Let z be any constant and define

$$G(u, z, S) := z + \sum_{i \in S_+} \mathbb{E}[(U_i - z)^+].$$

It is easy to see that $G(u, z, S)$ is an upper bound on $G(u, S)$ for all z . Let

$$\bar{G}(u, S) := \min_z G(u, z, S).$$

Since $G(u, z, S)$ is convex in z , the first order optimality condition is also sufficient and is given by:

$$\sum_{i \in S_+} \mathbb{P}(U_i \geq z) = 1.$$

The root of this equation, say $z^*(S)$, exists given our continuity assumptions. Consequently, $\bar{G}(u, S) = G(u, z^*(S), S)$.

We can approximate $\pi_i(S)$ by the gradient of $\bar{G}(u, S)$, which we denote by $\bar{\pi}$. Clearly

$$\bar{\pi}_i(S) = \frac{\partial \bar{G}(u, S)}{\partial u_i} = \mathbb{P}(U_i \geq z^*(S)) \quad i \in S_+.$$

Notice that $\bar{\pi}_i(S)$ depends only on the marginal distribution of ϵ_i and $z^*(S)$, which makes $\bar{\pi}_i(S)$ much easier to compute than $\pi_i(S)$.

Example 4.2 Suppose $U_i = u_i + \epsilon_i$ and that $\epsilon_i = \tau_i - \theta$ for all $i \in N_+$, where the τ_i 's are exponential random variables with mean θ . Let $v_i := \exp(u_i/\theta)$ for all $i \in N_+$. It is easy to see that in this setting,

$$\bar{\pi}_i(S) = \frac{v_i}{v_0 + V(S)} \quad \forall i \in S_+,$$

resulting in a BAM. Notice that the independence of the ϵ_i 's was not required.

Example 4.3 Let $U_i := u_i + \epsilon_i$ and $\epsilon_i := \theta - \tau_i$, where the τ_i 's are exponential with mean θ . Let $v_i := \exp(-u_i/\theta)$ for all $i \in N_+$ (notice the negative sign in the exponent). Suppose $S \subseteq N$ has the following property

$$|S| \max_{i \in S_+} v_i \leq v_0 + V(S) \quad \forall i \in S_+, \tag{4.10}$$

then a closed-form approximation for the exponential model described in Sect. 4.8 is given by:

$$\bar{\pi}_i(S) = \frac{v_0 + \sum_{j \in S} (v_j - v_i)}{v_0 + V(S)} \quad \forall i \in S_+, \quad (4.11)$$

where again the independence of the ϵ_i s was not required.

Condition (4.10) guarantees that all of the probabilities are non-negative. Otherwise the product with the largest v_i , $i \in S_+$ needs to be dropped. By repeatedly removing the product with the largest v_i from S_+ , we eventually obtain a new set S satisfying condition (4.10) so that (4.11) is a closed-form approximation to the exponential model. Notice that in the process of removing products, it is possible that product 0 is removed and in this case the term v_0 must be dropped from (4.11).

4.12 Choice Models and Retailing

Most of the applications of discrete choice models are to situations where the consumer would normally select a single alternative. Examples include transportation and lodging. As choice modeling and assortment optimization are finding their way into the retailing and e-commerce, the assumption of purchasing at most one product needs to be revisited as people may end up buying more than one pair of shoes even they go shopping with the expectation of buying a single pair. A reasonable model in the retail setting may be such that a customer selects non-negative thresholds z_i , $i \in S$, with the idea of buying all products whose utilities exceed their corresponding thresholds. Under this setting, the consumer may wish to maximize $\sum_{i \in S} \mathbb{E}[U_i | U_i \geq z_i] \mathbb{P}(U_i \geq z_i)$ subject to a bound, say $\sum_{i \in S} \mathbb{P}(U_i \geq z_i) \leq \kappa$, on the expected number of products to be purchased. We call this the threshold utility model (TUM). The upper bound $\bar{G}(u, S)$ on the consumer surplus $G(u, S)$, corresponds to the optimal selection of a uniform threshold ($z_i = z$ for all $i \in S$) for the case $\kappa = 1$ when the non-negativity of the thresholds is relaxed.

We can interpret the model in two different ways. First, after the consumer observes the utilities of the products, the consumer purchases all the products whose utility exceeds $z^*(S)$. The number of purchased products is random, and its expectation is κ . In this case $\bar{G}(u, S)$ measures directly the expected surplus. Alternatively, we can view the TUM as a consideration set model, where the consumer first observes the products $i \in S$ with $U_i \geq z^*(S)$, and then selects the one with the largest utility. If the set is empty, then the consumer selects the no-purchase alternative. It can be shown that under mild conditions the expected consumer surplus under this consideration set model is at least $e^{-\kappa}$ of $G(u, S)$.

4.13 End of Chapter Questions

1. Show that the IDM is a RUM by modeling it as a mixture of BAM's.
2. Show that the BAM and the GAM can be represented as rank-one MC choice models.
3. Give an interpretation of the general rank-one MC choice model.
4. Show that the random consideration set model can be represented as a MC choice model with a rank-one triangular transition matrix.
5. Consider a nested logit model with two nests. The assortment for nest 1, say S_1 , must be selected from the set $N_1 = \{1, 2\}$ and the assortment for nest two, say S_2 , must be selected from the set $N_2 = \{3, 4\}$. Suppose that dissimilarity parameters of the nests are $\gamma_1 = 0.8$ and $\gamma_2 = 0.5$. Finally, assume that $v_0 = 1, v_1 = 0.5, v_2 = 1.2, v_3 = 0.8$, and $v_4 = 0.5$.
 - (a) Compute the first choice probabilities for all $j = 0, 1, \dots, 4$.
 - (b) Compute the transition probabilities ρ_{ij} for all $i \neq j, i \in N = N_1 \cup N_2$ and $j \in N_+ = N \cup \{0\}$.
 - (c) Use the MC choice model to compute the probability that product 1 is selected if $S_1 = \{1\}$ and $S_2 = \{3, 4\}$ and compare this to the actual probabilities from the nested logit model.
 - (d) Repeat Part c for $S_2 = \{3\}$ and for $S_2 = \{4\}$.
6. Compute $G(u, S)$ for the MNL model and verify that $\pi(S)$ is the gradient of $G(u, S)$ with respect to u .
7. Consider a Probit model where $U_i = u_i + \epsilon_i$ where the ϵ_i 's are independent, mean zero normal random variables. Suppose that $N = \{1, 2\}$, $u = (0, 1)$ and the ϵ_i 's have standard deviation equal to 2. Use simulation to compute $\pi_i(N)$ for $i \in N_+$ and then use the approximation of Sect. 4.11 to compute $\bar{\pi}_i(N)$ for $i \in N_+$. Compare your results.
8. Let $U_i = u_i + \epsilon_i$ for all $i \in N_+$. Suppose there is a partition of the products $N = \cup_{j \in M} N_j$ so that for each $i \in N_j$, ϵ_i is an independent exponential with parameter θ_j . In addition, suppose that ϵ_0 is exponential with parameter 1. Let $v_i = \exp(u_i/\theta_j)$ for all $i \in N_j$, and $v_0 = \exp(u_0)$. Suppose that an assortment $S = \cup_{j \in M} S_j$ is offered where $S_j \subseteq N_j$ for each j . Show that there is a z^* such that

$$\sum_{j \in M} V_j(S_j) \exp(-z/\theta_j) + v_0 \exp(-z) = 1$$

where $V_j(S_j) = \sum_{i \in S_j} v_i$. Let $\gamma_j = \exp(1 - z^*/\theta_j)$, and show that

$$\bar{\pi}_i(S) = \frac{v_i}{V_i(S_j)} \frac{\gamma_j V_i(S_j)}{v_0 + \sum_{k \in M} \gamma_j V_k(S_k)} \quad \forall i \in S_j, \forall j \in M.$$

9. Consider the choice model (4.11) under assumption (4.10). Assume that the v 's are normalized so that $v_0 + V(N) = 1$. Show that the MC approximation is given by $\lambda_i = \pi_i(N) = 1 - nv_i$ for all $i \in N_+$ and $\rho_{ij} = v_j/(1 - v_i)$ for all $i \in N, j \in N_+$, with $\rho_{0,0} = 1$. Show that the approximation is exact and therefore (4.11) is a special case of the MC choice model.

4.14 Bibliographic Remarks

The BAM was first proposed by Luce (1959), where the author postulated the two choice axioms discussed in this chapter. Luce (1977) also provides a discussion of the same axioms. The classical reference for the MNL model is McFadden (1980). The equivalence between random utility models and choice models that result from probability distributions over preference lists is due to Block (1974). The GAM was introduced by Gallego et al. (2015) by slightly modifying the Luce axioms. Early references on the NL model date back to Domencich and McFadden (1975) and McFadden (1978). Huh and Li (2015) study a multi-stage version of the NL model, where the products are divided into nests, the nests are divided into subsets and so on. Koppelman and Wen (2000) examine the properties of the paired combinatorial logit (PCL) model, which is a version of the NL model. In the PCL model, each nest has at most two products, but a product can occur in multiple nests. Wen and Koppelman (2001) study the generalized NL model, which includes the NL and PCL models. McFadden and Train (2000) show that any RUM model can be approximated to any degree of accuracy by using a mixture of BAM's. The generalized NL model is a special case of a substantially broader class of choice models, called the generalized extreme value models, which are discussed in Train (2002). Wang (2018a) studies a variant of the BAM, where the attraction value of a product can depend on the size of the market it garners.

Manzini and Mariotti (2014) discuss maximum random consideration set models. Blanchet et al. (2016) introduce the MC choice model. The MC choice model has been shown to be a special case of the RUM by Berbeglia (2016). The exponential model was first introduced by Daganzo (1979) and further analyzed by Alptekinoglu and Semple (2016). Economists have been active developing rational inattention models, where consumers have a prior on the utility of each product and can reduce their uncertainty at a cost that is proportional to the reduction in entropy between the prior and the posterior. The resulting model has some semblance to the mixture of BAM's. The reader is referred to Caplin and Dean (2014) and Matejka and McKay (2015) for further information about rational inattention models. Ben-Akiva and Lerman (1985) give a thorough discussion of classical discrete choice models. The paper by van Ryzin (2005) discusses how discrete choice models can replace relying on the assumption that each consumer arrives with a fixed product in mind.

Huettner et al. (2018) study choice models where the consumers make a rational choice on which products to focus. Jagabathula and Rusmevichientong (2019) develop approaches to understand to what extent the choices of the consumers in the

data can be explained by using the random utility maximization principle. Natarajan et al. (2009), Mishra et al. (2012, 2014), and Ahipasaoglu et al. (2019) build choice models by assuming that moment or marginal distribution information about the utilities are available and they estimate the joint utility distribution as the one that maximizes the expected utility of the consumer from her most preferred option. The authors show that this estimation problem can be solved efficiently. Feng et al. (2017, 2018) show that other arguments can be used to construct choice models and some of these arguments can yield broader classes of choice models. Berbeglia (2018) presents the generalized stochastic preference model that contains interesting examples of discrete choice models that are not RUMs. The upper bound $\bar{G}(u, S)$ is based on the well-known Lai and Robbins (1976) bound for maximally dependent random variables, while the interpretation of the bound in the retail setting is due to Gallego and Wang (2019).

While estimation is not covered in detailed in this book, it is of crucial importance to the implementation of good revenue management solutions. Important papers on estimation of choice models include Berbeglia et al. (2018), Farias et al. (2013), Jagabathula and Vulcano (2018), Jagabathula et al. (2018), Kok and Fisher (2007), Martinez-de-Albeniz and Saez-de-Tejada (2014), Phillips et al. (2015, 2016), Simsek and Topaloglu (2018), van Ryzin and Vulcano (2015), and Vulcano and van Ryzin (2012). Some of these papers are based on adaptations of the EM method for parametric models, while others attempt to fit non-parametric models which have important advantages in capturing complex behavior that may be at odds with the parametric models covered in this book.