IE4214 REVENUE MANAGEMENT

CHARLES BAUER

1. Discrete Choice Models

In a Discrete Choice Model, a customer is presented with a finite number of alternatives labelled $N = \{1, ..., n\}$. The choice probability or attraction q_i is the probability that a customer chooses alternative $i \in N$.

In a Random Utility Model, each customer has random utility for alternative i given by $U_i = \mu_i + \epsilon_i$ where μ_i is deterministic and identical for all customers and ϵ_i is a random variable. The choice probability becomes $q_i = \mathbb{P}(i = \text{argmax}_{k \in N}(\mu_k + \epsilon_k))$.

1.0.1. *Example.* The utility for three products is $\mu = (1, 4, 7)$. The random component in customer utility, ϵ_k, is 4 with probability 0.5, 0 with probability 0.2, and -4 with probability 0.3 for all three products. Find q_1, the probability the first product is chosen.

Solution.

$$\begin{aligned} q_1 &= \mathbb{P}(i = \text{argmax}_{k \in N}(\mu_k + \epsilon_k)) \\ &= \mathbb{P}(\mu_1 + \epsilon_1 > \mu_2 + \epsilon_2, \mu_1 + \epsilon_1 > \mu_3 + \epsilon_3) \\ &= \mathbb{P}(\epsilon_1 = 4, \epsilon_2 = -4, \epsilon_3 = -4) + \mathbb{P}(\epsilon_1 = 4, \epsilon_2 = 0, \epsilon_3 = -4) \\ &= 0.5 * 0.3 * 0.3 + 0.5 * 0.2 * 0.3 = 0.075 \end{aligned}$$

1.1. **Multinomial Logit Model.** In the Multinomial Logit Model, $\epsilon_i \sim \text{Gumbel}(0,1)\ \forall i$. Thus, ϵ_i has probability distribution function (pdf) $f(x) = e^{-x}e^{-e^{-x}}$ and cummulative distribution function (cdf) $F(x) = e^{-e^{-x}}$.

The choice probability (as a function of the centers vector μ) follows.

$$\begin{aligned} q_i(\mu) &= \mathbb{P}(U_i \geq U_j \forall j \neq i) \\ &= \int_{-\infty}^{\infty} \mathbb{P}(\mu_i + \epsilon_i \geq \mu_j + \epsilon_j \forall j \neq i \big| \epsilon_i = x) f(x) dx \end{aligned}$$

$$\mathbb{P}(\mu_i + \epsilon_i \geq \mu_j + \epsilon_j \forall j \neq i \big| \epsilon_i = x) = \prod_{j \neq i} \mathbb{P}(\mu_i + x \geq \mu_j + \epsilon_j)$$

Date: February 2023.

$$
\begin{aligned}
&= \prod_{j \neq i} \mathbb{P}(\epsilon_j \leq \mu_i + x - \mu_j) \\
&= \prod_{j \neq i} F(\mu_i + x - \mu_j) \\
&= \prod_{j \neq i} e^{-e^{-(\mu_i + x - \mu_j)}} \\
&= e^{-\sum_{j \neq i} e^{-\mu_i - x + \mu_j}} \\
&= e^{-e^{-x} \sum_{j \neq i} e^{\mu_j - \mu_i}}
\end{aligned}
$$

$$
\begin{aligned}
q_i(\mu) &= \int_{-\infty}^{\infty} e^{-e^{-x} \sum_{j \neq i} e^{\mu_j - \mu_i}} e^{-x} e^{-e^{-x}} dx \\
&= \int_{x=-\infty}^{x=\infty} e^{-e^{-x} (\sum_{j \neq i} e^{\mu_j - \mu_i} + 1)} e^{-x} dx
\end{aligned}
$$

Let $a = \sum_{j \neq i} e^{\mu_j - \mu_i} + 1$, $y = e^{-x}$, and $dy = -e^{-x} dx$. Note $x = -\infty$ corresponds to $y = e^{-(-\infty)} = \infty$ and $x = \infty$ corresponds to $y = e^{-\infty} = 0$.

$$
\begin{aligned}
q_i(\mu) &= \int_{y=\infty}^{y=0} e^{-ya} (-1) dy \\
&= \frac{e^{-ya}}{a} \Big|_{y=\infty}^{y=0} \\
&= \frac{1}{a} \\
&= \frac{1}{\sum_{j \neq i} e^{\mu_j - \mu_i} + 1}
\end{aligned}
$$

Since $1 = e^0 = e^{\mu_i - \mu_i}$,

$$
q_i(\mu) = \frac{1}{\sum_{j \in N} e^{\mu_j - \mu_i}}
$$

$$
q_i(\mu) = \frac{e^{\mu_i}}{\sum_{j \in N} e^{\mu_j}} = \text{softmax}(\mu)_i = \sigma(\mu)_i
$$

1.1.1. *Practice.* Derive q_i for the Exponomial Model when $\epsilon_i \sim -\text{Exponential}(1)$.

1.1.2. *Independence of Irrelevant Alternatives.* The comparison of alternatives i, j only depends on the other centers μ_i, μ_j and not other alternatives. The odds ratio demonstrates this principle:

$$
\frac{q_i(\mu)}{q_j(\mu)} = \frac{\sigma(\mu)_i}{\sigma(\mu)_j} = \frac{e^{\mu_i}}{e^{\mu_j}} = e^{\mu_i - \mu_j}
$$

The principle implies adding one alternative decreases the choice probabilities for all other alternatives proportionately.

1.1.3. *Multinomial Probit Model.* What if $\epsilon \sim \text{Normal}(0, \Sigma)$? The variability in customer utilities would often be expected to follow a symmetric Normal distribution rather than the left-skewed Gumbel. The Multinomial Probit Model uses this assumption. However, the resulting choice probability does not have a closed-form and would be computed through Monte Carlo simulation. Optimization is more difficult.

1.1.4. *Subset of Choices.* Sometimes, not all alternatives are available. If a subset $S \in N$ is available, the choice probabilities become

$$q_{i,S} = \begin{cases} \frac{e^{\mu_i}}{\sum_{j \in S} e^{\mu_j}} & \forall i \in S \\ 0 & \text{otherwise} \end{cases}$$

Generally for a Random Utility Model, define a new utility vector $\hat{\mu}$ where $\hat{\mu}_j = \mu_j$ for $j \in S$ and $\hat{\mu}_j = -\infty$ otherwise. $q_{i,S} = q_i(\hat{\mu})$.

1.2. **Nested Logit Model.** The Nested Logit Model considers that adding an alternative has a greater effect on the choice probabilities of similar alternatives. The model uses a two-stage choice system where n alternatives are partitioned into K nests. An alternative is chosen by first choosing its corresponding nest and then choosing the specific alternative within the nest. For example, a customer first chooses either taxi or bus and then chooses either the red or blue bus. Let B_k be the set of alternatives in nest k. Additionally, select dissimilarity parameters $\lambda_k \in (0, 1]\ \forall k$.

The choice probability of alternative i belonging to nest k follows.

$$q_i(\mu) = \mathbb{P}(\text{nest } k \text{ chosen}) * \mathbb{P}(\text{alternative } i \text{ chosen} \mid \text{nest } k \text{ chosen})$$

$$q_i(\mu) = \frac{(\sum_{j \in B_k} e^{\mu_j/\lambda_k})^{\lambda_k}}{\sum_{l=1}^{k} (\sum_{j \in B_l} e^{\mu_j/\lambda_l})^{\lambda_l}} * \frac{e^{\mu_j/\lambda_k}}{\sum_{j \in B_k} e^{\mu_j/\lambda_k}}$$

1.2.1. *Example.* Alternatives $\{1, 2, 3, 4\}$ belong to nest 1, alternatives $\{5, 6\}$ belong to nest 2, and alternative 7 belongs to nest 3. $\mu_i = 0\ \forall i$, $\lambda_1 = 0.5$, $\lambda_2 = 1$, and $\lambda_3 = 0.7$. Find q_1, q_5, and q_7.

Solution. Since $\mu_i = 0\ \forall i$, $e^{\mu_j/\lambda_k} = 1\ \forall j, k$.

$$q_1 = \frac{(1+1+1+1)^{\lambda_1}}{(1+1+1+1)^{\lambda_1} + (1+1)^{\lambda_2} + 1^{\lambda_3}} * \frac{1}{4} = \frac{4^{0.5}}{4^{0.5} + 2^1 + 1^{0.7}} * \frac{1}{4} = 0.1$$

$$q_5 = \frac{2^{\lambda_2}}{4^{\lambda_1} + 2^{\lambda_2} + 1} * \frac{1}{4} = \frac{2^1}{4^{0.5} + 2^1 + 1} * \frac{1}{4} = 0.1$$

$$q_7 = 1 - 4q_1 - 2q_5 = 0.4$$

1.2.2. *What is λ?* Let's visualize with a simple example where alternatives 1,2 are in nest A and alternative 3 is in nest B. If $\lambda_A = 1$, $q_1 = \frac{2}{2^1+1}\frac{1}{2} = \frac{1}{3}$. Each alternative has the same probability independent of nest. As $\lambda_A \to 0$, $q_1 + q_2 \to \frac{1}{2^0+1} = \frac{1}{2}$. Each nest has the same probability. Accordingly, a lower lambda for a nest means the component alternatives are more similar and compete with each other more to reduce their individual choice probability when the others are present.

1.3. **Feature-based Choice Models.** The customers' utilities are no longer determined independently for each alternative. Now the utilities are determined by the features of an alternative, represented by the vector x. Assume μ_i is a linear regression of x: $\mu_i = \beta^T x$. Then for Multinomial Logit Model, $q_i = \text{softmax}(\beta^T x)_i$.

The parameters β can be trained from a dataset to result in the desired choice probabilities. Each set of alternatives present in the train data can be written as S_k The proportions of observations of choice i when presented with alternative set S_k can be written as y_{i,S_k}. The optimal β has maximum loglikelihood given the training data.

$$\hat{\beta} = \max_{\beta} \sum_{k=1}^{K} \sum_{i \in S_k \cup \{0\}} y_{i,S_k} \log q_{i,S_k}(\beta) = \max_{\beta} \sum_{k=1}^{K} \sum_{i \in S_k \cup \{0\}} y_{i,S_k} \log \text{softmax}(\beta^T x)_i$$

The maximum could be found in non-concave situations using stochastic gradient descent. Categorical features of n categories can be included as n indicator variables within x that correspond to distinct β.

1.3.1. *Mixed Logit Model.* In a homogenous model, β is the same for all customers and alternatives. A useful heterogenous model would cluster the customers into several groups, based on separate demographic data that details features assigned to customers, and calculate a β for each group.

If the probability of each group is known, The expected value of the choice probability can represent the probability for a customer of unknown group. Then β for a customer becomes a random variable with known distribution $g(\beta)$.

$$q_i = \begin{cases} \dfrac{\exp(\beta^T x_i)}{\sum_{j=1}^{n} \exp(\beta^T x_j)} g(\beta) & \text{for discrete } x \\ \displaystyle\int_{\beta} \dfrac{\exp(\beta^T x_i)}{\sum_{j=1}^{n} \exp(\beta^T x_j)} g(\beta) d\beta & \text{for continuous } x \end{cases}$$

In economics, $g(\beta)$ is often assumed to be Normal. Then there is no closed-form integral. A general form would be fit to the data through simulation or evaluation at specific values could be computed using an integral calculator. In revenue management, $g(\beta)$ is often discrete.

1.4. **Rank List Model.** Non-parametric, not RUM. Each customer has a rank list σ and the customer chooses the most preferred alternative among the offer set. For example, if a customer has $\sigma = \{1 \succ 3 \succ 2\}$ and $S = \{2, 3\}$ is offered, the customer will choose 3. Suppose the probability of each rank list is P_σ. Then $q_{i,S}$, the probability of a choice i when offered choices S, is simply the sum of the P_σ for σ such that i ranks first among S.

The number of possible rank lists is $n!$ so there are $n!$ parameters. Typically, P_σ is sparse i.e. $P_\sigma = 0$ for many σ.

1.4.1. *Theorem. Block and Marschak 1959.* $q_{i,S}$ comes from a rank list model if and only if there exists a distribution of U such that $q_{i,S} = \mathbb{P}(U_i \geq U_j, \forall j \in S)$.

In other words, if the utility μ of a Random Utility Model is assumed to be constant for all i, the Random Utility Model produces the same result as a Rank List Model.