ORIE 4154: Pricing and Market Design **Fall 2026**

Lecture 16 — Customer Choice: Luce, MNL, and Random Utility

Lecturer: Sid Banerjee *Date: October 20, 2026 (draft)*

16.1 Why we need choice probabilities

In Lecture 15's spiral-down example, a customer who preferred the inexpensive ticket L could still buy the expensive ticket H when L was unavailable. Counting yesterday's H sales as a sample of an autonomous "demand for H" missed this substitution. The seller had changed the customer's options, and therefore changed the data it observed.

Let $N = \{1, \dots, n\}$ be the products, $S \subseteq N$ the *offer set* (or assortment), and 0 the outside option of buying nothing. Write

$$q_i(S) = \mathbb{P}(\text{a customer chooses } i \mid S \text{ is offered}), \qquad i \in S \cup \{0\}. \tag{16.1}$$

The dependence on S is the point: $q_H(\{H\})$ may exceed $q_H(\{H, L\})$. Our question is what assumptions make the entire family of probabilities $\{q_i(S)\}$ both behaviorally meaningful and usable in optimization.

Lecture 2 generated binary purchase probabilities from random reservation values and briefly introduced binary logit. We now allow several products, so that customers can substitute *between* products as well as between purchase and no purchase.

16.2 A general discrete-choice model

For every assortment S, a choice model specifies numbers satisfying

$$q_i(S) \geq 0, \qquad q_i(S) = 0 \quad (i \notin S \cup \{0\}), \qquad q_0(S) + \sum_{i \in S} q_i(S) = 1. \tag{16.2}$$

In particular $q_0(\varnothing) = 1$. These requirements alone do not determine how the probabilities change when the seller changes S.

Example 16.1 (Closing the cheap fare). *Suppose every customer prefers L to H to leaving. Then*

S	$q_H(S)$	$q_L(S)$	$q_0(S)$
$\{H, L\}$	0	1	0
$\{H\}$	1	0	0

The same customers generate zero H sales under the first offer and one H sale per customer under the second. Nothing about their underlying preferences changed.

If m independent customers each see the same assortment S, the expected number of sales of product i is $mq_i(S)$. The *purchase probability* is $1 - q_0(S) = \sum_{i \in S} q_i(S)$; conditional on a purchase, product i's share is $q_i(S)/(1 - q_0(S))$, when the denominator is positive. These are different objects. In particular, conditioning on a purchase discards the possibility that a new product expands the market by drawing customers away from 0.

16.3 Regularity: a minimal substitution principle

Definition 1 (Regularity of choice). *A choice model is* regular *if, whenever $S \subseteq T$ and $i \in S \cup \{0\}$,*

$$q_i(T) \leq q_i(S). \tag{16.3}$$

Adding options cannot make an already available option more likely to be chosen. The outside option is included.

Regularity is a statement about substitution, not about the concavity of a revenue curve. Lecture 2 used the word *regular* for a distribution whose virtual value is increasing; that is a distinct definition.

Why is (16.3) natural? In a random-utility model, each customer draws utilities $U_0, U_1, \ldots, U_n$ and chooses the highest-utility offered alternative. Fix one realization of these utilities and use the same tie-breaking rule for every assortment. If i wins from T, then it also wins from the smaller set S containing i: removing competitors cannot make i lose. Thus

$$\mathbf{1}\{i \text{ chosen from } T\} \leq \mathbf{1}\{i \text{ chosen from } S\}.$$

Taking expectations proves regularity. The proof applies to $i = 0$ too. It requires a common underlying utility ranking (including consistent treatment of ties), rather than independently redrawing tastes for each offer set.

Regularity is weak. It restricts the *direction* of substitution, but not its size or destination. It will be the only behavioral assumption in Lecture 17's general-choice approximation theorem.

16.4 Luce's choice axiom and odds invariance

To state Luce's choice axiom (LCA) in its standard form, temporarily allow any nonempty *menu* $A \subseteq N \cup \{0\}$, including menus without 0. Let $\pi(i \mid A)$ denote the probability of choosing i when exactly A is available, with probabilities summing to one. Our course notation is the specialization

$$q_i(S) = \pi(i \mid S \cup \{0\}).$$

For a subset $C \subseteq A$, write $\pi(C \mid A) = \sum_{i \in C} \pi(i \mid A)$. Throughout this section assume strictly positive probabilities for every available alternative. We use the positive-probability form of Luce's axiom; his original formulation also addresses zero probabilities.

Definition 2 (Luce's choice axiom). *For $C \subseteq A \subseteq B$ with $A \neq \varnothing$,*

$$\boxed{\pi(C \mid B) = \pi(C \mid A)\,\pi(A \mid B).} \tag{16.4}$$

The probability of choosing from C under the larger menu B equals the probability of choosing from A under B, multiplied by the probability of choosing from C when only A is available. This relates behavior under *different menus*; it is an assumption about choice, rather than an automatic consequence of conditional probability. In particular, restricting the menu to A renormalizes the probabilities of its alternatives by a common factor.

16.4.1 Deriving independence of irrelevant alternatives

Take distinct $i, j \in B$ and apply LCA with $A = \{i, j\}$, first to $C = \{i\}$ and then to $C = \{j\}$:

$$\pi(i \mid B) = \pi(i \mid \{i,j\})\,\pi(\{i,j\} \mid B),$$
$$\pi(j \mid B) = \pi(j \mid \{i,j\})\,\pi(\{i,j\} \mid B).$$

Dividing cancels the common positive factor. Thus the odds of i versus j are the same in every menu containing them. Returning to offer sets S, T, we obtain *independence of irrelevant alternatives* (IIA):

$$\frac{q_i(S)}{q_j(S)} = \frac{q_i(T)}{q_j(T)}. \tag{16.5}$$

Here i, j belong to both $S \cup \{0\}$ and $T \cup \{0\}$, so j may be the outside option. The relative odds do not depend on which other products are offered. This restriction is much stronger than regularity.

For instance, suppose $q_H(\{H\}) = 2/3$ and $q_0(\{H\}) = 1/3$. The probabilities $q_H(\{H, L\}) = 1/3$, $q_0(\{H, L\}) = 1/6$, and $q_L(\{H, L\}) = 1/2$ retain the $2:1$ odds of H over no purchase, even though both probabilities fall. IIA requires this invariance, not merely $q_H(\{H, L\}) \leq q_H(\{H\})$.

Theorem 16.2 (Luce representation). *Assume positive choice probabilities on every nonempty menu and LCA (16.4). Then there are positive weights $v_0, v_1, \ldots, v_n$ such that*

$$q_i(S) = \frac{v_i}{v_0 + \sum_{j \in S} v_j} \quad (i \in S), \qquad q_0(S) = \frac{v_0}{v_0 + \sum_{j \in S} v_j}. \tag{16.6}$$

Conversely, the corresponding probabilities $\pi(i \mid A) = v_i / \sum_{j \in A} v_j$ satisfy LCA, IIA, and regularity.

Proof: Use the outside option as a common reference. Set $v_0 = 1$ and, for each product, define

$$v_i = \frac{q_i(\{i\})}{q_0(\{i\})} > 0.$$

For any S containing i, apply IIA to alternatives i and 0 in S and $\{i\}$:

$$\frac{q_i(S)}{q_0(S)} = \frac{q_i(\{i\})}{q_0(\{i\})} = v_i.$$

Hence $q_i(S) = v_i q_0(S)$. Normalization in (16.2) gives

$$1 = q_0(S) + \sum_{i \in S} q_i(S) = q_0(S) \left(1 + \sum_{i \in S} v_i\right).$$

Solving for $q_0(S)$ and substituting back yields (16.6). Multiplying every v_i, including v_0, by the same positive constant changes no probability, so $v_0 = 1$ is only a normalization.

Conversely, the weight formula on arbitrary menus gives, for $C \subseteq A \subseteq B$ with A nonempty,

$$\pi(C \mid A)\, \pi(A \mid B) = \frac{\sum_{i \in C} v_i}{\sum_{j \in A} v_j} \frac{\sum_{j \in A} v_j}{\sum_{k \in B} v_k} = \pi(C \mid B).$$

Thus LCA holds. The odds are v_i / v_j, independent of the menu, and enlarging a menu increases the denominator, proving regularity. ■

Formula (16.6) is the *multinomial logit* (MNL) choice model. The v_i are *attraction weights*, not prices or probabilities. A large v_0 makes no purchase more likely. Only weight *ratios* are identified by choices. Zero-weight products can be handled as limiting or degenerate cases, but the theorem's clean positive-odds statement excludes them.

16.5 Why Gumbel random utility gives MNL

The axiom gives one route to MNL; random utility gives another. Suppose each alternative, including 0, has utility

$$U_i = u_i + G_i,$$

where the G_i are independent, identically distributed standard type-I extreme-value (Gumbel) shocks. The customer chooses the available alternative with the largest U_i.

Here is a short derivation of the Gumbel-max identity. If E_i are iid exponential random variables of rate 1, then $G_i = -\log E_i$ has Gumbel CDF $\mathbb{P}(G_i \le x) = \exp(-e^{-x})$. Maximizing

$$u_i - \log E_i = \log\left(\frac{e^{u_i}}{E_i}\right)$$

is equivalent to minimizing E_i / e^{u_i}. These scaled variables are independent exponentials with rates e^{u_i}. In an exponential race, the chance that i finishes first is its rate divided by the sum of rates. Therefore

$$q_i(S) = \frac{e^{u_i}}{e^{u_0} + \sum_{j \in S} e^{u_j}}, \tag{16.7}$$

which is (16.6) with $v_i = e^{u_i}$. Adding the same constant to all u_i has no effect: only utility differences matter. The common shock scale has been normalized to one; changing it rescales the u_i.

The iid assumption does real work. It makes the relative odds between two products independent of what else is offered. Correlated tastes or heterogeneous customer segments can break that property while preserving regularity.

16.5.1 The same choice probabilities from an entropy regularizer

The exponential race also identifies the *value* of the noisy maximum. Fix $A = S \cup \{0\}$ and let $z = \sum_{i \in A} e^{u_i}$. The minimum race time

$$Y = \min_{i \in A} \frac{E_i}{e^{u_i}}$$

is exponential of rate z, so zY is exponential of rate 1. Consequently,

$$M := \max_{i \in A}(u_i + G_i) = -\log Y = \log z - \log(zY).$$

The last term is a standard Gumbel random variable. Writing $\gamma \approx 0.5772$ for its mean gives $\mathbb{E}[M] = \log z + \gamma$.

Now choose a probability vector $x = (x_i)_{i \in A}$ instead of a single alternative. Its mean deterministic utility is $\sum_i x_i u_i$, and its *entropy* is

$$h(x) = -\sum_{i \in A} x_i \log x_i, \qquad 0 \log 0 := 0.$$

Entropy rewards spreading probability across alternatives. The connection to Gumbel maximization is

$$\boxed{\mathbb{E}[M] - \gamma = \log z = \max_{\substack{x_i \geq 0 \ (i \in A) \\ \sum_{i \in A} x_i = 1}} \left\{ \sum_{i \in A} x_i u_i + h(x) \right\}.} \tag{16.8}$$

To derive the optimization identity, use concavity of log (Jensen's inequality):

$$\begin{aligned} \sum_{i \in A} x_i u_i + h(x) &= \sum_{i : x_i > 0} x_i \log \frac{e^{u_i}}{x_i} \\ &\leq \log \left(\sum_{i : x_i > 0} e^{u_i} \right) \leq \log z. \end{aligned}$$

Both inequalities are equalities exactly when $x_i = e^{u_i}/z$ for every $i \in A$. Thus the unique optimizing probabilities are precisely the MNL probabilities $q_i(S)$ from the Gumbel derivation.

Without the entropy term, maximizing $\sum_i x_i u_i$ puts all probability on alternatives with the largest u_i. With noise scale $\tau > 0$, utilities $u_i + \tau G_i$ instead give the regularizer $\tau h(x)$ and probabilities proportional to $e^{u_i/\tau}$. The noise scale therefore controls how strongly the choice is smoothed away from a hard maximum.

16.6 The red-bus/blue-bus test

Suppose a customer chooses between a car and a red bus, each with attraction weight 1. Treat the car as the reference alternative in this illustration. MNL predicts probability 1/2 for each. Now add a blue bus that is essentially a duplicate of the red one and give it the same weight. MNL predicts

$$\mathbb{P}(\text{car}) = \tfrac{1}{3}, \qquad \mathbb{P}(\text{red bus}) = \tfrac{1}{3}, \qquad \mathbb{P}(\text{blue bus}) = \tfrac{1}{3}.$$

It assigns the two buses a total probability of 2/3. If customers instead view the buses as near-perfect substitutes, a plausible prediction is to keep car at 1/2 and split the original bus probability: 1/4 red, 1/4 blue. MNL cannot produce that pattern while retaining the same car-to-red-bus odds; those odds must stay $1 : 1$.

MNL *does* satisfy regularity: adding blue bus lowers the probabilities of both car and red bus. The alternative $1/2, 1/4, 1/4$ pattern is regular too, but it breaks IIA. The difficulty is *proportional substitution*, or global IIA, not a failure of regularity. When new products are close substitutes for particular existing products, this restriction can be consequential for pricing and assortment decisions.

16.7 A brief look beyond MNL

Nested logit.
Group similar products into nests. A dissimilarity parameter permits stronger within-nest substitution, relaxing global IIA; ordinary nested logit is not universal.

Mixed MNL.
Average MNL probabilities across customer segments with different attraction vectors. Rich mixtures can approximate broad random-utility patterns under suitable conditions, not all stochastic choice rules.

Rankings and Markov-chain choice.
A distribution over rankings or a Markov chain for substitution can capture patterns a single MNL misses.

Flexibility can make assortment optimization harder. Lecture 17 compares exact MNL revenue ordering with a regularity-only approximation.

16.8 Takeaways and next question

1. A choice model describes $q_i(S)$ for each *offer set*, not an autonomous demand number for each product.
2. Random utility implies regularity: adding alternatives cannot help an existing one. This is not Lecture 2's virtual-value definition of a regular distribution.
3. Luce's multiplicative choice axiom implies odds invariance (IIA), which leads to MNL using the outside option as a reference.
4. Maximization with iid Gumbel shocks and optimization with an entropy regularizer produce the same MNL probabilities.
5. MNL is tractable but can misdescribe substitution among similar products.

The choice model tells us what happens *if* we offer S. The seller's next decision is which S to offer. Is a larger assortment always better for revenue?

Recommended reading

- **Core:** Talluri and van Ryzin, *The Theory and Practice of Revenue Management*, Section 7.2.2 (random utility and MNL).
- **Luce's axiom:** R. Duncan Luce, "Luce's choice axiom", *Scholarpedia*, 2008. This explains the original formulation from *Individual Choice Behavior* (1959).
- **Optional optimization connection:** Boyd and Vandenberghe, *Convex Optimization*, Example 3.25, for the log-sum-exp/entropy relationship.
- **Course connection:** Lectures 2 (binary logit) and 15 (spiral down).
- **Deeper companion:** Banerjee, ORIE 6154 Choice Models notes.