ECOLE POLYTECHNIQUE

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

ORDERED CHOICE PROBABILITIES IN RANDOM UTILITY MODELS

André de PALMA

Karim KILANI

March 1, 2015

Cahier n° 2015-04

DEPARTEMENT D'ECONOMIE

Route de Saclay

91128 PALAISEAU CEDEX

(33) 1 69333033

http://www.economie.polytechnique.edu/

mailto:chantal.poujouly@polytechnique.edu

Ordered choice probabilities in random utility models*

André de Palma[†] Karim Kilani[‡]

March 1, 2015

Abstract

We prove a general identity which states that any element of a tuple (ordered set) can be obtained as an alternating binomial weighted sum of first elements of some sub-tuples. The identity is then applied within the random utility models framework where any alternative's ordered choice probability (the probability that it has a given rank) is expressed with respect to standard best choice probabilities. The logit and the logsum formulas are extended to their ordered choice counterparts. In a symmetric case, we compare for the probit and the logit, the surplus loss due to the withdrawal of a product with the damage due to the loss of a rank. *Keywords*: Generalized Roy's identity; Logit; Ordered utilities; Order statistics; Probit; Random utility models. *JEL classification*: D11

*The authors are grateful to participants in the lunch seminar of Ecole Polytechnique, France, in the International Choice Modeling Conference 2011, in the Kuhmo Nectar Conference 2014, in the seminars in National Taiwan University, Taiwan, in ETH, Zurich, Switzerland. We owe special thanks to Moshe Ben-Akiva, Gilbert Laffond, Jen Lainé, and Claude Lefèvre.

†ENS Cachan, Centre d'Economie de la Sorbonne, Cachan, France; CECO ’ Ecole Polytechnique, France. Tel.: +33-6-63-64-4320; E-mail: andre.depalma@ens-cachan.fr

‡CNAM, LIRSA, Paris 75003, France. Tel.: +33-1-40-27-2366; E-mail: karim.kilani@cnam.fr

1 Introduction

Structural demand models have been very much used to estimate disaggregated data collected at the individual level. Probabilistic choice models, and in particular behavioral choice models (such as the logit, nested logit, mixed logit, probit, etc.) play here an important role. They are known as discrete choice models (DCM), which are very much used in many fields of economics and beyond (trade, industrial organization, labor economics, demography, urban and transport economics, marketing, etc. (see McFadden (2001)). These tools describe and evaluate the behavior of individuals selecting their best alternative from a discrete set (usually finite and known) of alternatives (see Anderson, de Palma, and Thisse (1992)).

Most researchers in DCM and in particular in random utility models (RUM) focused their attention on the choice of the best alternative. The selected option, or best alternative, corresponds to the one yielding the highest realized utility. Theoretical formula for the probability of the best choice are crucial for the modeler mainly for building demand models, to study the theoretical properties of the demand system; simulation of choice probabilities are needed for estimating the parameters entering into the utility (relying on standard econometric package). The ranking of preferences has attracted much less attention.

However, ranks matter. (1) Quite often in surveys, elicitation of preference is achieved via stated preference questions, i.e. by asking the respondent to rank alternatives, such as by Netflix. (2) Alternatively, Amazon, tour operators, doctors, health satisfaction surveys, investment advisers do not suggest the best alternative, but an array of ordered alternatives (see Greene and Hensher (2010)). (3) For a variety of reasons, the agents may not choose their best alternative since it may not be available (or known) when the decision was made (moreover, how much is lost in this case?). (4) Lack of full rationality or lack of full information provide other justifications for studying lower-order choices.

We believe that rankings are more likely to matter with increasingly complex technologies, larger variety of goods and services, which make individual choices for the best alternative more difficult. A recent body of the literature has made significant progresses in the direction of integrating lower-order or sub-optimal choices (see e.g. Berry, Levinsohn, and Pakes (2004)). These authors use first and second-best choices revealed by respondents to survey questionnaires. They show that it is possible to obtain much better estimates in the context of the choice of differentiated products and to reduce the burden associated to surveys. In the marketing literature, Marley and Louviere (2005) consider the extreme situation where respondents are asked to reveal both their best and their worse preferred alternative. They provide new insights on model parameter estimations obtained from this type of data.

Researchers who wish to incorporate more than just the optimal alternative in their study are interested to know how to derive expressions for the probabilities of full rank-ordered choice profiles, where the alternatives are ordered from the best choice to the worst one. Unfortunately, this problem is difficult and explicit expressions exist in very few cases, such as for the rank-ordered logit (see Beggs, Cardell, and Hausman (1981)). However, even when such expressions exist (as in the logit case) they do not allow to recover simple expressions for single lower-order choices. This is mainly due to intricate combinatorics inherent to these problems.

Some researchers have added new assumptions to the main axioms of standard discrete choice approach, to get for example, expressions for the worst choice probabilities (see Marley and Louviere (2005)). Other approaches, such as the ordered probit model, which use threshold value, has also been proposed. Ordered choice modeling, mixed and hierarchical models include different ways to introduce thresholds, which represent one natural manner to model ranking probabilities. Finally, an alternative solution is the use of EM algorithms, Monte-Carlo simulation techniques exploiting Gibbs sampler, allowed by faster computer's processors in order to deal with lower-order choice models (see Train (2003)). This solution presents an interest for estimation, but is bounded to be slow for large choice sets, and it remains less attractive from a theoretical point of view.

In this paper, we adopt a very conservative view. We wish to keep the standard discrete choice random utility approach and study the ranking probabilities and welfare of lower ranked alternatives. The simple analysis of the logit model suggests some research agenda. It shows that the probabilities of ranks can be expressed as a sum of products of standard logit models. However, these expressions seem very much related to the strong hypothesis behind the logit (i.i.d. and Gumbel error terms). Extension to the probit, even with i.i.d. normal distribution appears to be an impossible task. Moreover, the standard logit analysis remains totally silent about the welfare properties of lower ranks, while the welfare of the best choice is well known since 40 years ago (see the derivation of the nested logit and logsum by Ben-Akiva (1973)).

The literature remains silent a fortiori for any other RUM. These questions will be addressed in this paper. Moreover, the relation between choice probabilities and welfare given by Roy's identity, is shown to remain true for lower ranks. Clearly, these results have to make abstraction of the error terms underlying the construction of the RUM. We proceed more generally with a system of choice probabilities. A rather long detour to order statistics will be required. The main mathematical background beyond this paper is drawn from order statistics. In our setting, we focus on the r-th order statistics, i.e. on r-th highest value (standard order-statistics consider the k-th lowest values). We focus on the relation with the maximum (the reasoning for the minimum would be the same).

The maximum plays a key role in economics, since the maximum of random variables corresponds to the alternative chosen by the individual in a RUM, and since the expected maximum utility is the consumer surplus (when there are no income effects). We start with a deterministic approach, with ordered sets. We wish to compare elements in different order and in different ordered sets. To fixed idea consider an example. Assume that a student is ranked 5-th in a class of 20 students. S/he should be encouraged if 6 months later, s/he turns out to be 4-th out of 20. However, s/he should also be encouraged if the size of the class increases from 20 to 25 students, while maintaining the 5-th rank. However, it is harder to decide whether or not s/he should be more encouraged in the first situation than in the second.

To address this question, we start with a deterministic version of the triangular inequalities, used in order statistics with random variables (see e.g., Arnold, Balakrishnan, and Nagaraja (2008)). Doing that, we are able to recover a deterministic version of the triangular inequality removing the restrictions on the correlation patterns. This approach will play a key role in this paper to obtain results valid for random utility model with any correlation patterns. Our first result (Lemma 1) relates the gain from higher rank with the gain from the same rank in larger set. This identity relates ranks and size. Such result is standard for random variables in order statistics (triangular identity). Here it is shown to be valid for: (a) any element (not only for random variables but also for real numbers); and (b) set of elements not necessarily ordered. In this case, elements in the choice set are denoted by labels (which play the role of the rank).

This deterministic triangular identity (Lemma 1) corresponds to a recurrence which can be solved. The main result of this paper (Theorem 2) provides the element of order r in a set C as a function the highest order element in different subsets of C appropriately chosen. Then, we show (Theorem 3) that the same identity is valid for any operator $\mathbf{T}(\cdot)$. When $\mathbf{T}(\cdot)$ is the expectation operator, and when elements are random utilities, Corollary 6 shows that any expectation of the r-th ordered utility can be expressed as expectations of maximum utility within appropriately chosen subsets. Similarly, Theorem 6 allows to compute the probability of rank r as a function of the standard (best) choice probabilities. We then provide a new version of Roy's identity, which is valid for any rank, and not, as in the standard case, for the best choice probabilities. To sum-up, our setting allows to generate the following economic results, valid for any random utility models (with the nested logit and the GEV as special case): (1) Relation between ranking probabilities and maximum probabilities. (2) Welfare of the ordered choice as a function of best-choice expected utility. (3) Relation between ranking probabilities and welfare of ordered choices, extending Roy's identity to ordered random utility models.

The paper is organized as follows. Section 2 introduces the setting for tuples of ordered objects and derived our ground results, Theorem 2, where an identity is obtained. Section 3 applies these results in the RUM framework where standard results appearing in the order statistics literature are derived as corollaries of Theorem 3. We also derive a relation between expectation of order utilities and expectation of maximum utilities. Choice probabilities in the RUM context are introduced in Section 4 where a new identity relating any order choice probability to best choice probabilities is obtained. It allows to derive a generalization of Roy's identity, valid for any welfare of expected ranked utility (Theorem 8). Application to the standard multinomial logit, to GEV models, and an illustrative example in the symmetric case are discussed in Section 5.

2 Preliminaries results

2.1 The main theorem

Consider an n-tuple (ordered set, list) $l_n \equiv (l_n(1), \ldots, l_n(n))$ of n elements of a vector space (real or complex numbers, real matrices, random variables, etc.), $n \geq 2$. Let $l_p \equiv (l_p(1), \ldots, l_p(p))$ denotes a p-tuple of p *different* elements drawn from l_n, $1 \leq p \leq n$, assuming that the order of appearance of the elements in l_p remains the same as in l_n. We denote by $\mathcal{L}_p$ the set of such p-tuples, whose cardinality is $\binom{n}{p}$.

Note that any l_{n-1} of $\mathcal{L}_{n-1}$ can be obtained by dropping a single element $l_n(k)$ from l_n, $1 \leq k \leq n$. For convenience, such $(n-1)$-tuple is instead denoted by $l_n^{[k]}$, with r-th order element $l_n^{[k]}(r)$, $1 \leq r \leq n-1$. Our first result is a technical lemma.

Lemma 1. *The following recurrence relation holds:*

$$\sum_{k=1}^{n} l_n^{[k]}(r) = r l_n(r+1) + (n-r) l_n(r), \ 1 \leq r \leq n-1. \tag{2.1}$$

Proof. When building $l_n^{[k]}$, if the dropped element from l_n has a rank k in l_n such that $k \leq r$ (resp. $k > r$), $1 \leq r \leq n-1$, the r-th order element of $l_n^{[k]}$ matches the $(r+1)$-th (resp. r-th) order element of l_n. This observation yields the following equation

$$l_n^{[k]}(r) = \mathbf{1}_{\{k \leq r\}} l_n(r+1) + \mathbf{1}_{\{k > r\}} l_n(r),$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator operator. Summing up over k both sides of the above equation, we obtain

$$\sum_{k=1}^{n} l_n^{[k]}(r) = \left(\sum_{k=1}^{n} \mathbf{1}_{\{k \leq r\}}\right) l_n(r+1) + \left(\sum_{k=1}^{n} \mathbf{1}_{\{k > r\}}\right) l_n(r).$$

Using the fact that: $\sum_{k=1}^{n} \mathbf{1}_{\{k \leq r\}} = n - \sum_{k=1}^{n} \mathbf{1}_{\{k > r\}} = r$, Eq. (2.1) is obtained. □

In what follows, $a_q^m \equiv (-1)^{m-q} \binom{m}{q}$ will denote *alternating binomial coefficients*, where m and q are two integers such that $0 \leq q \leq m$. The cornerstone result of this paper is now stated.

Theorem 2. *The following identity, relating the r-th order element of the tuple to first elements of some of its sub-tuples, holds:*

$$l_n(r) = \sum_{p=n-r+1}^{n} a_{n-r}^{p-1} \sum_{l_p \in \mathcal{L}_p} l_p(1), \; 1 \le r \le n. \tag{2.2}$$

Proof. Identity (2.2) states that $l_n(r) = R_n(r)$, where $R_n(r)$ is the the RHS of (2.2): $R_n(r) \equiv \sum_{p=n-r+1}^{n} a_{n-r}^{p-1} \sum_{l_p \in \mathcal{L}_p} l_p(1)$. For any $n \ge 2$ and $r = 1$, the sum contains only one term. Using the fact that $a_{n-1}^{n-1} = 1$ and that $\mathcal{L}_n$ contains the single tuple l_n, we obtain $l_n(1) = R_n(1)$. For $r = 2$, the first sum runs from $p = n-1$ to $p = n$. Since $a_{n-2}^{n-1} = -(n-1)$, we get

$$R_n(2) = \sum_{k=1}^{n} l_n^{[k]}(1) - (n-1)\, l_n(1).$$

Eq. (2.1) with $r = 1$ implies $\sum_{k=1}^{n} l_n^{[k]}(1) = l_n(2) + (n-1)\, l_n(1)$, so that the above equation yields $l_n(2) = R_n(2)$. Therefore, (2.2) is true for any $n \ge 2$ and for $r = 1, 2$. For $n = 2$, note that (2.2) is true for all the possible ranks ($r = 1, 2$). For $n \ge 3$, by induction, assume that (2.2) is valid for any $(n-1)$-tuple and all possible rank r verifying $1 \le r \le (n-1)$. This implies that, for any k such that $1 \le k \le n$, we have

$$l_n^{[k]}(r) = \sum_{p=n-r}^{n-1} a_{n-1-r}^{p-1} \sum_{l_p \in \mathcal{L}_p^{[k]}} l_p(1),$$

where $\mathcal{L}_p^{[k]}$ is the set of all p-tuples, $1 \le p \le n-1$, drawn from $l_n^{[k]}$ in the same manner as the tuples drawn from l_n, i.e. by keeping the same order for the selected elements. We prove that (2.2) is true for n and any possible rank r veryfing $1 \le r \le n$. As seen above, (2.2) is true for $r = 1, 2$. By induction on r, assume that $R_n(r-1) = l_n(r-1)$ for a given rank r, $1 \le r \le n-1$. The tuples of $\mathcal{L}_p^{[k]}$ consists of tuples of $\mathcal{L}_p$ but those containing the element $l_n(k)$, their set being denoted hereafter by $\mathcal{L}_p^{\leftarrow k}$, must be subtracted. This observation allows us to rewrite the last above equation as

$$l_n^{[k]}(r) = \sum_{p=n-r}^{n-1} a_{n-1-r}^{p-1} \left[\sum_{l_p \in \mathcal{L}_p} l_p(1) - \sum_{l_p \in \mathcal{L}_p^{\leftarrow k}} l_p(1) \right].$$

By summation over all k, we get

$$\sum_{k=1}^{n} l_n^{[k]}(r) = \sum_{p=n-r}^{n-1} a_{n-1-r}^{p-1} \left[n \sum_{l_p \in \mathcal{L}_p} l_p(1) - \sum_{k=1}^{n} \sum_{l_p \in \mathcal{L}_p^{\leftarrow k}} l_p(1) \right].$$

Within the double summation appearing into the brackets, any tuple l_p of $\mathcal{L}_p$, $n-r \leq p \leq n-1$, is accounted for as many times as its number of elements, i.e. p times. Consequently, the following expression is obtained

$$\sum_{k=1}^{n} l_n^{[k]}(r) = \sum_{p=n-r}^{n-1} (n-p)\, a_{n-1-r}^{p-1} \sum_{l_p \in \mathcal{L}_p} l_p(1).$$

Note that the summation of the RHS can be extended to n, obtaining

$$\sum_{k=1}^{n} l_n^{[k]}(r) = \sum_{p=n-r}^{n} (n-p)\, a_{n-1-r}^{p-1} \sum_{l_p \in \mathcal{L}_p} l_p(1).$$

It is worth to rewrite it as

$$\sum_{k=1}^{n} l_n^{[k]}(r) = rR_n(r+1) + \sum_{p=n-r}^{n} (n-p-r)\, a_{n-1-r}^{p-1} \sum_{l_p \in \mathcal{L}_p} l_p(1).$$

Note that for $p = n-r$, the corresponding term of the sum of the RHS is null, allowing to rewrite the above expression as

$$\sum_{k=1}^{n} l_n^{[k]}(r) = rR_n(r+1) + \sum_{p=n-r+1}^{n} (n-p-r)\, a_{n-1-r}^{p-1} \sum_{l_p \in \mathcal{L}_p} l_p(1).$$

Using an elementary property of binomial coefficients, the last equation becomes

$$\sum_{k=1}^{n} l_n^{[k]}(r) = rR_n(r+1) + (n-r) \sum_{p=n-r+1}^{n} a_{n-r}^{p-1} \sum_{l_p \in \mathcal{L}_p} l_p(1),$$

yielding

$$\sum_{k=1}^{n} l_n^{[k]}(r) = rR_n(r+1) + (n-r) R_n(r).$$

By induction, we have assumed that $R_n(r) = l_n(r)$, so that

$$\sum_{k=1}^{n} l_n^{[k]}(r) = rR_n(r+1) + (n-r) l_n(r).$$

Comparing the above equation with Eq. (2.1), we get $l_n(r+1) = R_n(r+1)$.

□

The above theorem, which provides an identity relating tuple's elements to first order elements of sub-tuples, is now applied in a RUM framework. We derive new identities among ordered utilities, their distribution, their expectation, etc., and then derive an identity for ordered choice probabilities.

3 Application to random utility models

There are $n \geq 2$ objects of choice referred by their "labels". The set of n labels of the objects of choice (alternatives) is denoted by C. In a RUM framework, the utility U_i derived by an individual selecting the object with label i (alternative i), $i \in C$, is modeled as a real-valued *random variable* defined on a common probability space $(\Omega, \mathcal{F}, P)$.

For one particular realization $\{U_i(\omega)\}_{i \in C}$, $\omega \in \Omega$, of the finite set of *random utilities* $\{U_i\}_{i \in C}$, the alternatives referred by their label, can be rank-ordered according to a decreasing order of magnitude of their associated utilities. The top rank alternative has the highest utility while the bottom rank alternative has the lowest utility. When two or more alternatives have the same utility, the order among them can be chosen arbitrary, e.g. using their alphabetic order.

Denote by $(U_{1:n}(\omega), ..., U_{n:n}(\omega))$ the n-tuple of "ordered utilities", where $U_{1:n}(\omega) \equiv \max_{i\in C} U_i(\omega)$, ..., $U_{n:n}(\omega) \equiv \min_{i\in C} U_i(\omega)$. We refer to $U_{r:n}$, the random variable with realizations $U_{r:n}(\omega)$, $\omega \in \Omega$, as the r-th *ordered utility*, $1 \le r \le n$.

The results which follows are the most general and use operators denoted by $\mathbf{T}(\cdot)$ which are mappings from the vector space of real random variables defined on $(\Omega, \mathcal{F}, P)$ into another vector space. This general setting allows us to apply Theorem 2 to ordered utilities, to their CDFs, their expectations, etc.

Theorem 3. *For any operator* $\mathbf{T}(\cdot)$ *mapping the space of real random variables into another vector space, the following identity holds among ordered utilities:*

$$\mathbf{T}(U_{r:n}) = \sum_{p=n-r+1}^{n} a_{n-r}^{p-1} \sum_{A_p \subseteq C} \mathbf{T}\left(\max_{k\in A_p} U_k\right), \ 1 \le r \le n, \tag{3.1}$$

where A_p denotes subsets of C with p cardinality.

Proof. Define the n-tuple $l_n \equiv (\mathbf{T}(U_{1:n}), \cdots, \mathbf{T}(U_{n:n}))$, where its r-th order element is $\mathbf{T}(U_{r:n})$, $1 \le r \le n$. Thanks to Theorem 2, we have

$$\mathbf{T}(U_{r:n}) = \sum_{p=n-r+1}^{n} a_{n-r}^{p-1} \sum_{l_p \in \mathcal{L}_p} l_p(1),$$

where $l_p(1)$ is the first element of a p-tuple l_p drawn from l_n. Let $\mathcal{L}_p(1) \equiv \{l_p(1) : \ l_p \in \mathcal{L}_p\}$ be the set of elements of l_n appearing as first elements of some $l_p \in \mathcal{L}_p$. Note that $\mathcal{L}_p(1)$ has $\binom{n}{p}$ elements. Recall that the ordered utilities used in l_n verify: $U_{1:n} \geq \cdots \geq U_{n:n}$. Therefore,

$$\mathcal{L}_p(1) = \left\{ \mathbf{T}\left(\max_{k \in A_p} U_k \right) : A_p \subseteq C \right\}.$$

Note that there are $\binom{n}{p}$ subsets A_p. Therefore, the sum $\sum_{l_p \in \mathcal{L}_p} l_p$ coincides with $\sum_{A_p \subseteq C} \mathbf{T}\left(\max_{k \in A_p} U_k\right)$. Putting together the above observations, the required equation is obtained. □

The above theorem can be applied to derive an identity among ordered utilities.

Corollary 4. *The ordered utilities verify the following identity:*

$$U_{r:n} = \sum_{p=n-r+1}^{n} a_{n-r}^{p-1} \sum_{A_p \subseteq C} \max_{k \in A_p} U_k, \; 1 \leq r \leq n. \tag{3.2}$$

Proof. Apply Theorem 3 using as $\mathbf{T}(\cdot)$ operator the identity of the space of real random variables defined on $(\Omega, \mathcal{F}, P)$, to obtain the required identity. □

Eq. (2.1) implies:

$$\frac{\sum_{k \in C} \left(U_{r:n} - U_{r:n}^{[k]} \right)}{r} = U_{r:n} - U_{r+1:n} =, \; 1 \leq r \leq n-1, \tag{3.3}$$

where $U_{r:n}^{[k]}$ denotes the r-th order utility when alternative k is dropped from the choice set C. The LHS of (3.3) is an *average* utility gain due to a larger choice. It coincides with the RHS which measures the utility gain derived from a better ranking.

Recurrence relations and identities appearing in the order statistics literature apply to the distributions of the order statistics rather than the order statistics themselves (or their realizations). In this vein, let $F_i(x) \equiv P(U_i \leq x)$, $x \in \mathbb{R}$, be the marginal CDF of U_i, $i \in C$. Under assumptions kept the most general for now, we show that a standard result of the order statistics literature can be seen as a corollary of Theorem 3.

Corollary 5. *The CDFs of the ordered utilities verify, for any real x, the following identity:*

$$P\left(U_{r:n} \leq x\right) = \sum_{p=n-r+1}^{n} a_{n-r}^{p-1} \sum_{A_p \subseteq C} P\left(\max_{k \in A_p} U_k \leq x\right), \ 1 \leq r \leq n. \tag{3.4}$$

Proof. Let $\mathbf{T}(\cdot)$ be the following operator $\mathbf{T}(X) = P(X \leq x)$, where X is any random variables defined on $(\Omega, \mathcal{F}, P)$. It maps the space of real random variables into $[0,1]$. Application of Theorem (3) with such operator yields Eq. (3.4). □

We now apply Theorem 3 to derive an identity among the expectations of the ordered utilities.

Corollary 6. The expectations of *the ordered utilities,* $\mu_{r:n} \equiv \mathbf{E}(U_{r:n})$, when they exist, verify *the following identity:*

$$\mu_{r:n} = \sum_{p=n-r+1}^{n} a_{n-r}^{p-1} \sum_{A_p \subseteq C} \mathbf{E}\left(\max_{k \in A_p} U_k\right), \ 1 \leq r \leq n. \tag{3.5}$$

Proof. We apply Theorem 3 with $\mathbf{T}(\cdot)$ chosen as being the expectation operator: $\mathbf{T}(X) = \mathbf{E}(X)$, which maps the space of real random variables into the real line. □

Note that no linearity is required for the $\mathbf{T}(\cdot)$ operator. Similar corollaries can also be stated for any m-th order moment (taking $\mathbf{T}(X) = \mathbf{E}(X^m)$), or even for variances of the order statistics (taking $\mathbf{T}(X) = \mathbf{Var}(X)$, where $\mathbf{Var}(\cdot)$ is the variance operator).

4 Ordered choice probabilities

4.1 Choice probabilities

Assumptions about the probability distribution of the random utilities $\{U_i\}_{i \in C}$ are slightly strengthened. Its CDF given by $P(U_i \leq x_i,\ i \in C)$ is assumed to be absolutely continuous with respect to Lebesgue measure over $\mathbb{R}^n$. One consequence of absolute continuity is that ties among utilities occur with zero probability, so that almost surely, there is a *strict ranking* among the alternative utilities.

Consider the event $(U_{r:n} = U_i)$ where it occurs that $i \in C$ corresponds to the r-th order choice in the choice set C, $1 \leq r \leq n$. Order choice probabilities $\mathbf{P}_{r:n}(i)$ are defined, for $i \in C$, as being the probability of this event $\mathbf{P}_{r:n}(i) \equiv P(U_{r:n} = U_i)$. The sequence (indexed by the alternatives) of events $\{(U_{r:n} = U_i)\}_{i \in C}$ form a partition of Ω (up to a null-measure set), implying that: $\sum_{i \in C} \mathbf{P}_{r:n}(i) = 1$, $1 \leq r \leq n$. The same property prevails for the sequence (indexed by the ranks) of events $\{(U_{r:n} = U_i)\}_{1 \leq r \leq n}$, so that: $\sum_{r=1}^{n} \mathbf{P}_{r:n}(i) = 1$, $i \in C$.

An important topic in RUMs is the derivation of an expression for the best choice probabilities $\mathbf{P}_{A_p}(i) \equiv P\left(\max_{k \in A_p} U_k = U_i\right)$, $i \in A_p$, $A_p \subseteq C$. Recall they can be obtained by performing the following integration (see e.g. Anderson, de Palma, and Thisse (1992))

$$\mathbf{P}_{A_p}(i) = \int_{-\infty}^{+\infty} \frac{\partial P(U_i \leq x_i,\ i \in A_p)}{\partial x_i}\Big|_{x_i = x} dx, \tag{4.1}$$

where the symbol $\big|_{x_i = x}$ means that the partial derivative has to be taken at $x_i = x$, $i \in A_p$.

In the logit case where alternative utilities are independent Gumbels, a closed form recalled in Section 5.1 is obtained. This is not usually the case for any RUM, as for the probit where utilities are normally distributed, requiring simulation techniques exposed in Train (2003), in order to compute the integral (4.1).

4.2 Identity for choice probabilities

Theorem 3 is now adapted to derive a new identity for order choice probabilities.

Theorem 7. *For any RUM, for any* alternative $i \in C$*, the ordered choice probabilities are related to best choice probabilities by:*

$$\mathbf{P}_{r:n}(i) = \sum_{p=n-r+1}^{n} a_{n-r}^{p-1} \sum_{\{i\} \subseteq A_p \subseteq C} \mathbf{P}_{A_p}(i), \; i \in C, \; 1 \leq r \leq n. \tag{4.2}$$

Proof. Consider the operator given by $\mathbf{T}(X) = P(X = U_i)$, mapping the set of random variables X defined on $(\Omega, \mathcal{F}, P)$ into $[0, 1]$. Application of Theorem 3 with such operator yields

$$P(U_{r:n} = U_i) = \sum_{p=n-r+1}^{n} a_{n-r}^{p-1} \sum_{A_p \subseteq C} P\left(\max_{k \in A_p} U_k = U_i\right).$$

For any subset A_p verifying $i \notin A_p$, thanks to the absolute continuity of the distribution of the utilities $\{U_i\}_{i \in C}$, we have $P\left(\max_{k \in A_p} U_k = U_i\right) = 0$, implying the required identity. □

4.3 Williams-Daly-Zachary theorem extended

Additive random utility models (ARUMs) assume that utilities have an additive form: $U_i = v_i + \varepsilon_i$, $i \in C$, where v_i is the systematic part of the utility and ε_i is the random error term. Utilities are assumed here to have finite expectation.

The Williams-Daly-Zachary theorem (see e.g. McFadden (1981)) states that the derivative with respect to v_i of the expected maximum utility (first-order moment) within a choice subset A_p, allows to recover the best choice probabilities

$$\frac{\partial \mathbf{E}\left(\max_{k\in A_p} U_k\right)}{\partial v_i} = \mathbf{P}_{A_p}\left(i\right),\ i \in A_p. \tag{4.3}$$

More generally, we prove that the derivation of the expected order utilities $\mu_{r:n}$ with respect to v_i allows to recover the order choice probabilities.

Theorem 8. *For any ARUM where utilities have finite expectation, we have:*

$$\frac{\partial \mu_{r:n}}{\partial v_i} = \mathbf{P}_{r:n}\left(i\right),\ i \in C,\ 1 \leq r \leq n. \tag{4.4}$$

Proof. Deriving Eq. (3.5), with respect to v_i, we obtain

$$\frac{\partial \mu_{r:n}}{\partial v_i} = \frac{\partial}{\partial v_i} \sum_{p=n-r+1}^{n} a_{n-r}^{p-1} \sum_{A_p \subseteq C} \mathbf{E}\left(\max_{k\in A_p} U_k\right).$$

Inverting the derivation and the sum signs of the RHS, we get

$$\frac{\partial \mu_{r:n}}{\partial v_i} = \sum_{p=n-r+1}^{n} a_{n-r}^{p-1} \sum_{A_p \subseteq C} \frac{\partial \mathbf{E}\left(\max_{k\in A_p} U_k\right)}{\partial v_i}.$$

For any A_p such that $i \notin A_p$, since $\partial \mathbf{E}\left(\max_{k\in A_p} U_k\right) / \partial v_i = 0$, the above equation can be written as

$$\frac{\partial \mu_{r:n}}{\partial v_i} = \sum_{p=n-r+1}^{n} a_{n-r}^{p-1} \sum_{\{i\} \subseteq A_p \subseteq C} \frac{\partial \mathbf{E}\left(\max_{k\in A_p} U_k\right)}{\partial v_i}.$$

Then using Eq. (4.3), we get

$$\frac{\partial \mu_{r:n}}{\partial v_i} = \sum_{p=n-r+1}^{n} a_{n-r}^{p-1} \sum_{\{i\} \subseteq A_p \subseteq C} \mathbf{P}_{A_p}\left(i\right).$$

Thanks to Identity (4.2), the RHS of the above equation is $\mathbf{P}_{r:n}(i)$, obtaining the required identity. □

5 Applications

5.1 The multinomial logit model

The multinomial logit model (MNL) is an ARUM with independent Gumbel distributed utilities with marginal CDFs given, for any real x, by

$$P(U_i \leq x_i) = \exp\left(-e^{v_i - x_i - \gamma}\right),\ i \in C, \tag{5.1}$$

where γ is Euler's constant. Note that v_i is the expectation of the utility: $\mathbf{E}(U_i) = v_i$.

A closure property of the Gumbel distribution ensures that the CDF of $\max_{k \in A_p} U_k$, the maximum utility among the alternatives of A_p, remains a Gumbel with an expectation given by

$$\mathbf{E}\left(\max_{k \in A_p} U_k\right) = \ln \sum_{k \in A_p} e^{v_k},\ A_p \subseteq C, \tag{5.2}$$

the celebrated *logsum formula* often used as a welfare measure in empirical works.

Applying Identity (3.5) of Corollary 6 and using the logsum formula (5.2), a *generalized logsum formula* is obtained for the expected r-th order utility within the MNL framework

$$\mu_{r:n} = \sum_{p=n-r+1}^{n} a_{n-r}^{p-1} \sum_{A_p \subseteq C} \ln \sum_{k \in A_p} e^{v_k},\ 1 \leq r \leq n. \tag{5.3}$$

Recall that the best choice probabilities also have a closed form given, for

$i \in A_p$, by the *logit formula*

$$\mathbf{P}_{A_p}(i) = \frac{e^{v_i}}{\sum_{k \in A_p} e^{v_k}},\ i \in\ A_p. \tag{5.4}$$

Application of Eq. (4.2) of Theorem 7 combined with the above logit formula yields a *generalized logit formula* for r-th order choice probabilities

$$\mathbf{P}_{r:n}(i) = \sum_{p=n-r+1}^{n} a_{n-r}^{p-1} \sum_{\{i\} \subseteq A_p \subseteq C} \frac{e^{v_i}}{\sum_{k \in A_p} e^{v_k}},\ 1 \leq r \leq n. \tag{5.5}$$

For example, the second-best logsum is given by

$$\mu_{2:n} = \sum_{i \in C} \ln \sum_{k \in C-\{i\}} e^{v_k} - (n-1) \ln \sum_{k \in C} e^{v_k}, \tag{5.6}$$

while the second-best logit choice probabilities are given by

$$\mathbf{P}_{2:n}(i) = \sum_{j \in C-\{i\}} \frac{e^{v_i}}{\sum_{k \in C-\{j\}} e^{v_k}} - (n-1) \frac{e^{v_i}}{\sum_{k \in C} e^{v_k}}. \tag{5.7}$$

A factorization of the above expression yields an alternative form

$$\mathbf{P}_{2:n}(i) = \sum_{j \in C-\{i\}} \frac{e^{v_j}}{\sum_{k \in C} e^{v_k}} \frac{e^{v_i}}{\sum_{k \in C-\{j\}} e^{v_k}}. \tag{5.8}$$

In accordance with formulas of the rank-ordered logit (see Beggs, Cardell, and Hausman (1981)), every term of the sum of the RHS corresponds to the joint probability that j is the best choice and i is the second best in C.

On the other hand, the expected utility corresponding to the worst choice is given by

$$\mu_{n:n} = \sum_{p=1}^{n} (-1)^{p-1} \sum_{A_p \subseteq C} \ln \sum_{k \in A_p} e^{v_k}. \tag{5.9}$$

It provides a useful benchmark for welfare measures. Moreover, using (5.5), the worst choice probabilities where minimum utility is achieved can be written as an alternating sum of logit expressions

$$\mathbf{P}_{n:n}(i) = \sum_{p=1}^{n} (-1)^{p-1} \sum_{A_p \subseteq C} \frac{e^{v_i}}{\sum_{k \in A_p} e^{v_k}},\ i \in C. \tag{5.10}$$

This formula has been derived by de Palma, Kilani, and Laffond (2013) making use of the Inclusion-Exclusion principle in probability theory.

5.2 Correlation among utilities

The MNL can be extended to allow for correlation among the random utilities which are still assumed to have Gumbel margins with CDFs given by Eq. (5.1). Correlation is introduced via *absolutely continuous* copulas (for an introduction to copulas, see e.g. Nelsen (1999)) $\Theta(\cdot)$ which are CDF functions defined over the unit n-cube $[0,1]^n$. The multivariate CDF of the random utilities is becoming

$$P(U_i \leq x_i,\ i \in C) = \Theta\left(\exp\left(-e^{v_1 - x_1 - \gamma}\right), ..., \exp\left(-e^{v_n - x_n - \gamma}\right)\right). \tag{5.11}$$

We focus our attention to the class of *extreme value copulas* (EVC, also referred to as *max-stable copulas*), which verify the following property (see Salvadori et al. (2007), p. 192)

$$\Theta\left(\phi_1^{\lambda}, ..., \phi_n^{\lambda}\right) = C^{\lambda}(\phi_1, ..., \phi_n),\ \forall \lambda > 0. \tag{5.12}$$

Under the above assumptions, the maximum utility within a choice subset $A_p \subseteq C$ has a CDF which can be written as

$$P\left(\max_{k \in A_p} U_i \leq x\right) = \exp\left(-G_{A_p} e^{-x}\right), \tag{5.13}$$

where G_C (referred to as tail dependence functions) is defined as

$$G_C \equiv -\ln \Theta \left(\exp\left(-e^{v_1}\right), ..., \exp\left(-e^{v_n}\right)\right), \tag{5.14}$$

while any other G_{A_p} is obtained by setting $v_k = -\infty$ for all $k \notin A_p$ in the RHS of the above expression.

A consequence of Eq. (5.13) is that the maximum utility is also Gumbel distributed with expected value given by

$$\mathbf{E}\left(\max_{k \in A_p} U_k\right) = \ln G_{A_p}. \tag{5.15}$$

The above formula provides a generalization of the logsum formula (5.2) to the GEV framework. Thanks to the Williams-Daly-Zachary theorem, best choice probabilities can be derived using the following

$$\mathbf{P}_{A_p}(i) = \frac{\partial \ln G_{A_p}}{\partial v_i},\ i \in A_p. \tag{5.16}$$

They correspond to the GEV probabilities obtained by McFadden (1978).

Corollary 6 allows a generalization of (5.15) to any order choice

$$\mu_{r:n} = \sum_{p=n-r+1}^{n} a_{n-r}^{p-1} \sum_{A_p \subseteq C} \ln G_{A_p},\ 1 \leq r \leq n. \tag{5.17}$$

Moreover, thanks to Theorem 7 and using (5.16), the r-th order choice probabilities are given by

$$\mathbf{P}_{r:n}(i) = \sum_{p=n-r+1}^{n} a_{n-r}^{p-1} \sum_{\{i\} \subseteq A_p \subseteq C} \frac{\partial \ln G_{A_p}}{\partial v_i},\ i \in C,\ 1 \leq r \leq n \tag{5.18}$$

Example 1 *The Gumbel-Hougaard (or logistic) family copula, the only Archi-*

median family of max-stable copulas, has the following form

$$\Theta\left(\phi_1,...,\phi_{\mathfrak{c}}\right)=\exp\left\{-\left[\sum_{i\in C}\left(-\ln\phi_i\right)^{\theta}\right]^{\frac{1}{\theta}}\right\},\ \theta\geq 1. \tag{5.19}$$

Note that the boundary case $\theta=1$ *coincides with the product (independent) copula:* $\Theta\left(\phi_1,...,\phi_{\mathfrak{c}}\right)=\prod_{i\in\mathfrak{C}}\phi_i$*. The associated tail dependence functions are* $G_{A_p}=\left(\sum_{k\in A_p}e^{\theta v_k}\right)^{\frac{1}{\theta}}$*. Order choice probabilities for such model are given by*

$$\mathbf{P}_{r:n}\left(i\right)=\sum_{p=n-r+1}^{n}a_{n-r}^{p-1}\sum_{\{i\}\subseteq A_p\subseteq C}\frac{e^{\theta v_i}}{\sum_{k\in A_p}e^{\theta v_k}},\ i\in C,1\leq r\leq n. \tag{5.20}$$

As $\theta=1$*, they coincide with the logit order choice probabilities given by (5.5).*

Example 2 *Let* $\{C_1,...,C_m\}$ *be a partition of the choice set* C *into* m *subsets or groups of alternatives, and construct a mixture of a Gumbel-Hougaard copula given by (5.19) and the independent copula in the following way*

$$\Theta\left(\phi_1,...,\phi_{\mathfrak{c}}\right)=\exp\left\{-\left[\sum_{g=1}^{m}\left(-\sum_{i\in C_g}\ln\phi_i\right)^{\theta}\right]^{\frac{1}{\theta}}\right\},\ \theta\geq 1. \tag{5.21}$$

Tail dependence functions are given by $G_{A_p}=\left[\sum_{g=1}^{m}\left(\sum_{i\in A_p\cap C_g}e^{v_i}\right)^{\theta}\right]^{\frac{1}{\theta}}$*. Therefore, order choice probabilities of any alternative belonging to group* g *are given, for* $1\leq r\leq n$*, and* $i\in C_g$*,*

$$\mathbf{P}_{r:n}\left(i\right)=\sum_{p=n-r+1}^{n}a_{n-r}^{p-1}\sum_{\{i\}\subseteq A_p\subseteq C}\left(\frac{\sum_{k\in A_p\cap C_g}e^{v_k}}{G_{A_p}}\right)^{\theta}\frac{e^{v_i}}{\sum_{k\in A_p\cap C_g}e^{v_k}}, \tag{5.22}$$

generalizing to any order the nested logit best choice probabilities introduced *by Ben Akiva (1973).*

5.3 The symmetric case

In the symmetric case, arising for example with i.i.d. utilities, using (2.1), the following recursion rule is obtained

$$n\mu_{r:n-1} = r\mu_{r+1:n} + (n-r)\,\mu_{r:n},\ 1 \leq r \leq n-1. \tag{5.23}$$

Only the computation of the expected best utilities ($\mu_{1:2}, \ldots, \mu_{1:n}$) are needed, the remaining expectations are then computed iteratively from (5.23), which can also be written as

$$\frac{\mu_{r:n} - \mu_{r+1:n}}{\mu_{r:n} - \mu_{r:n-1}} = \frac{n}{r},\ 1 \leq r \leq n-1. \tag{5.24}$$

Hence, the expected damage due to a loss in one rank is higher than the one due to the loss of an alternative.

We consider a symmetric MNL where (utilities are centered) $\mu_{1:n} = \ln n$. We also consider the case of i.i.d. normal utilities with zero mean and variance $\pi^2/6$, in order to get the same variance as for the Gumbel case. Note that for the normal case, explicit values of $\mu_{1:n}$ can be obtained for $n \leq 5$.[1] For larger values of n, numerical integration is needed.

Expected order utilities are computed using a spreadsheet for the Gumbel case and the R statistical software for performing numerical integration for the normal case. The results are displayed in the two figures below. The upper curve represents the expected maximum utility, which is used to deduce the lower figures representing expected lower-order utilities.

[1]For the standard normal distribution, $\mu_{1:2} = \pi^{-1/2}$; $\mu_{1:3} = 1.5\pi^{-1/2}$; $\mu_{1:4} = 6\pi^{-3/2}\tan^{-1}\sqrt{2}$; $\mu_{1:5} = 15\pi^{-3/2}\tan^{-1}\sqrt{2} - 2.5\pi^{-1/2}$ (see Arnold et al. (2008)).

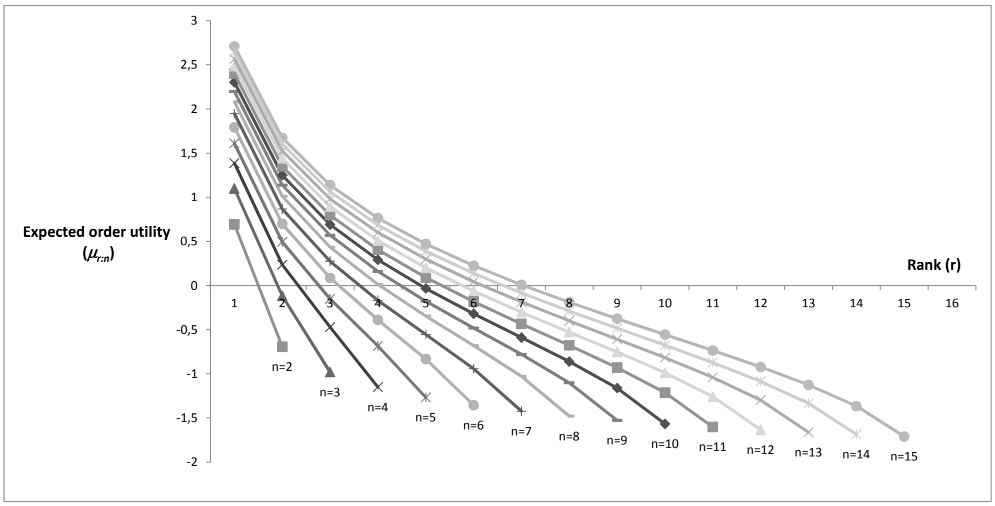

Figure 5.1: Expected utility vs. rank order choice: logit

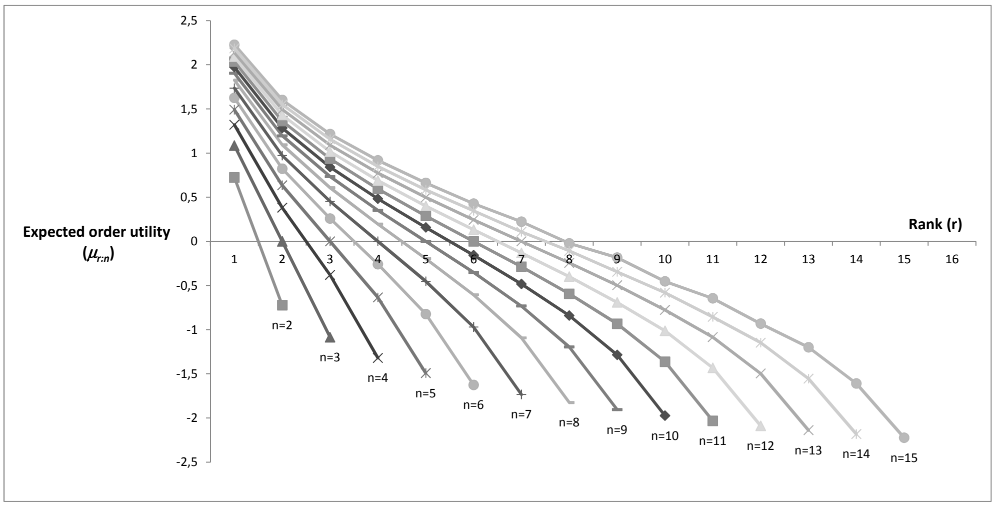

Figure 5.2: Expected utility vs. rank order choice: probit

The two figures look very similar. Several comments are in order. The expected maximum utility is increasing and concave while the expected minimum utility is decreasing and convex. For a given number of alternatives, the r-th order utility decreases with r and the utility loss when losing one rank is first decreasing and then increasing with the rank. This means that the loss from the penultimate to the last rank is significant. The expected utility of the r-th rank (computed from the top) increases with $\mathfrak{c}$ while the expected utility of the r'-th rank (computed from the bottom) decreases with n, consistently with the triangular condition (5.23). This means in particular that the penultimate for n alternatives is better than the penultimate for $n+1$ alternatives. Finally, about the Julius Caesar's quote "I had rather be the first in this village than second in Rome," we see that the answer depends on the relative sizes of Rome and the village. For example, the first among six alternatives is better than the second among fifteen.

References

[1] Anderson, S.P., de Palma, A., and Thisse, J.-F. (1992). *Discrete Choice Theory of Product Differentiation*. MIT Press, Cambridge, MA.

[2] Arnold, B.C., Balakrishnan, N., and Nagaraja, H.N. (2008). *A First Course in Order Statistics*. SIAM Publishers, Philadelphia.

[3] Beggs, S., Cardell, S., and Hausman, J. (1981). Assessing the Potential Demand for Electric Cars, *Journal of Econometrics*, 17(1), 1–19.

[4] Ben-Akiva, M.E. (1973). *Structure of Passenger Travel Demand Models*, PhD *thesis*, Department of Civil Engineering, MIT, Cambridge, Ma.

[5] Berry, S., Levinsohn, J., and Pakes, A. (2004). Differentiated products demand systems from a combination of micro and macro Data: The new car market, *Journal of Political Economy*, 112, 1, 68-105.

[6] Greene, W.H and Hensher, D.A. (2010) *Modeling Ordered Choices: A Primer*. Cambridge University Press, Cambridge.

[7] Marley, A.A.J., and Louviere, J.J. (2005) Some probabilistic models of best, worst, and best–worst choices, *Journal of Mathematical Psychology*, 49, 6, 464-480.

[8] McFadden, D. (1978). Modelling the choice of residential location, in A. Karlquist et al. (ed.), *Spatial Interaction Theory and Residential Location*, North-Holland, Amsterdam, pp. 75-96.

[9] McFadden, D. (1981). Econometric Models of Probabilistic Choice, in C.F. Manski and D. McFadden (eds.), *Structural analysis of discrete data with econometric applications*, 198-272, MIT Press: Cambridge, MA.

[10] McFadden, D. (2001). Economic choices, *American Economic Review*, 91, 3, 351-378.

[11] Nelsen, R.B. (1999), *An Introduction to Copulas*, Springer: New York.

[12] de Palma, A., Kilani, K., and Laffond, G. (2013). Best and worst choices, *Working Papers* halshs-00825656, HAL.

[13] Salvadori, G., De Michele, C., Kottegoda, N.T., Rosso, R. (2007) *Extremes in Nature: An Approach Using Copulas*, Springer: New York.

[14] Train, K. (2003). *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, MA.