Multiple Discrete Choice with Order Statistic Marginal Utilities

Scott Webster
scott.webster@asu.edu
W. P. Carey School of Business, Arizona State University
April 2, 2020

Author note to readers (July 17, 2020): The genesis of this paper came from a thought: Could order statistics of an extreme value distribution be useful for modeling how individuals choose the number of units to purchase (or select) among a set of discrete alternatives? After analyzing several versions of the model presented here, I concluded: *probably not*. The reason is that the model exhibits properties that may be too restrictive and insufficiently rich to capture key drivers of these decisions.

On the other hand, some of the properties exposed through analysis are curious. Furthermore, the resulting joint and marginal probability distribution functions that appear in this paper may have value in other contexts.

Please acknowledge this paper if you find any of the content useful for your research. Also, I am open to discussions on research ideas that may build on results in this paper.

The paper presents a random utility maximization model for individuals selecting discrete quantities from a set of n alternatives. Multiple alternatives with positive quantities may be selected. Diminishing marginal utility to quantity of each alternative is modeled via order statistics of independent Gumbel random variables. The model is parsimonious—requiring $n + 1$ parameters—and tractable—admitting closed-form expressions for choice probabilities. As such, the model is amenable to maximum likelihood estimation of structural parameters from observed choices.

Probability functions recover binary logit probabilities under binary choice and a maximum quantity of one unit, and probability is monotonic in the quantity of each alternative. The monotonic property likely restricts the application of the model to a narrow range of settings. The property is a manifestation of a simple recursive relationship among Gumbel order statistic probabilities. This relationship, to my knowledge, has not previously been identified in the literature and may lead to new models for capturing important complexities in a tractable manner.

Keywords: Statistics; Utility-Preference, Choice Functions

1. Multiple Discrete Choice / Quantity Models

1.1. General Model

Individual j from a population selects quantities among n alternatives to maximize utility. Let $u_{ij}(x_i)$ denote individual j utility from $x_i \geq 0$ units of alternative i. Function $u_{ij}(x_i)$ is concave and increasing for all i and j. Individual j's total utility from decision $\mathbf{x} = (x_1, \ldots, x_n)$ is

$$u_j(\mathbf{x}) = \sum_{i=1}^{n} u_{ij}(x_i),$$

i.e., utility is additively separable in the alternatives.

Each individual chooses $\mathbf{x}$ to maximize total utility subject to relevant limits, a problem that can be formulated in two equivalent forms:

$$\text{P1: } \max_{\mathbf{x}} \left\{ u_j(\mathbf{x}) : \sum_{i=1}^{n} x_{ij} \le B_j \right\} \tag{1}$$

$$\text{P2: } \max_{\mathbf{x}} \left\{ u_j(\mathbf{x}) : \Delta u_{ij}(x_i) \le \Delta u_{0j}, i = 1, \ldots, n \right\} \tag{2}$$

In P1, an individual maximizes total utility subject to a budget constraint. In P2, an individual maximizes total utility of up to the point where the marginal utility from each alternative is not more than the marginal utility from the best outside alternative, which is identified by subscript 0. In this formulation, $\Delta u_{ij}(x_i)$ is the marginal utility of alternative i and Δu_{0j} is the marginal utility of the best outside alternative. For example, if decision $\mathbf{x}$ is continuous and u_{ij} are differentiable, then

$$\Delta u_{ij}(x_i) = \frac{\partial u_{ij}(x_i)}{\partial x_i},$$

and if $\mathbf{x}$ is limited to nonnegative integers, then

$$\Delta u_{ij}(x_i) = u_{ij}(x_i) - u_{ij}(x_i - 1).$$

Due to concave increasing u_{ij}, P1 can be equivalently stated as

$$\text{P1: } \max_{\mathbf{x}} \left\{ u_j(\mathbf{x}) : \Delta u_{ij}(x_i) \le \lambda_j, i = 1, \ldots, n \right\} \tag{3}$$

where λ_j is the Lagrange multiplier at budget B_j, i.e., λ_j in P1 plays the role of Δu_{0j} in P2. If $\mathbf{x}$ is continuous, then at optimal solution $\mathbf{x}^*$,

$$\Delta u_{ij}(x_i^*) = \lambda_j \text{ for all } x_i^* > 0 \text{ and } \Delta u_{ij}(x_i^*) < \lambda_j \text{ for all } x_i^* = 0.$$

1.2. Models for Continuous Quantity Choices

Bhat (2005) proposes the multiple discrete-continuous extreme value (MDCEV) model that is based on formulation P1 with continuous $\mathbf{x}$. Because $\mathbf{x}$ is continuous and u_{ij} are continuous, increasing, differentiable functions, the budget constraint is binding, which implies

$$x_i > 0 \text{ for some } i \in \{1, \ldots, n\} \tag{4}$$

for each individual. This is not restrictive, e.g., if $\mathbf{x} = 0$ among alternatives 1 through n is possible, then an alternative 0 corresponding to not selecting 1 through n can be introduced to the choice set.

MDCEV is appealing because (1) it admits closed-form probabilities for choice decisions among a population of individuals and (2) it reduces to the widely accepted multinomial logit (MNL) choice probability if only one alternative is selected. I summarize this model below.

The utility of alternative i of a randomly selected individual from the population is

$$\tilde{u}_i(x_i) = e^{a_i + \tilde{\varepsilon}_i}(b_i + x_i)^{\alpha_i}$$

where $\tilde{\varepsilon}_i$ are iid Gumbel random variables that capture idiosyncratic (unknown to the researcher) preferences of individuals. The remaining parameters, a_i, b_i, α_i reflect observed characteristics of alternative i utility, and $\alpha_i \in (0, 1]$ to reflect positive and diminishing marginal utility to consumption.

Note that for realization ε_i of $\tilde{\varepsilon}_i$,

$$\frac{\partial u_i(x_i)}{\partial x_i} = e^{a_i + \varepsilon_i + \ln\left(\alpha_i (b_i + x_i)^{\alpha_i - 1}\right)}.$$

Assuming the same Lagrange multiplier λ among individuals in the population, optimal $\mathbf{x}^*$ satisfies

$$a_i + \varepsilon_i + \ln(\alpha_i) + (\alpha_i - 1)\ln(b_i + x_i^*) = \ln\lambda \text{ for all } x_i^* > 0$$

$$a_i + \varepsilon_i + \ln(\alpha_i) + (\alpha_i - 1)\ln(b_i) < \ln\lambda \text{ for all } x_i^* = 0.$$

Define

$$v_i(x_i) = a_i + \ln(\alpha_i) + (\alpha_i - 1)\ln(b_i + x_i)$$

$$v_0 = \ln\lambda$$

$\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_n)$ = quantity choice vector for a randomly selected individual from the population.

Then the fraction of the population that selects quantity vector $\mathbf{x} = (x_1, \ldots, x_k, 0, \ldots, 0)$ where $x_i > 0$ for $i = 1, \ldots, k$ and $k \in \{1, \ldots, n\}$ is

$$P[\tilde{\mathbf{x}} = \mathbf{x}] = P\left[v_1(x_1) + \tilde{\varepsilon}_1 = v_0, \ldots, v_k(x_k) + \tilde{\varepsilon}_k = v_0, v_{k+1}(0) + \tilde{\varepsilon}_{k+1} < v_0, \ldots, v_n(0) + \tilde{\varepsilon}_n < v_0\right].$$

(Due to (4), $P[\tilde{\mathbf{x}} = \mathbf{0}] = 0$.). Since $\tilde{\varepsilon}_i$ are iid, $P[\tilde{\mathbf{x}} = \mathbf{x}]$ can be expressed as a product of probabilities, i.e.,

$$P[\tilde{\mathbf{x}} = \mathbf{x}] = \left(\prod_{i=1}^{k} P\left[\tilde{\varepsilon}_i = v_0 - v_i(x_i)\right]\right)\left(\prod_{i=k+1}^{n} P\left[\tilde{\varepsilon}_i < v_0 - v_i(0)\right]\right).$$

Substituting Gumbel probabilities into the above, normalizing, and simplifying yields

$$P[\tilde{\mathbf{x}} = \mathbf{x}] = \left(\prod_{i=1}^{k} \frac{1 - \alpha_i}{b_i + x_i}\right)\left(\sum_{i=1}^{k} \frac{b_i + x_i}{1 - \alpha_i}\right)\left(\frac{\prod_{i=1}^{k} e^{v_i(x_i)}}{\left(\sum_{j=1}^{n} e^{v_j(x_j)}\right)^k}\right)(k-1)! \tag{5}$$

(Bhat 2005). If $k = 1$, then (5) reduces to the form of MNL choice probability, i.e.,

$$P[\tilde{\mathbf{x}} = (x_1, 0, \ldots, 0)] = \frac{e^{v_1(x_1)}}{\sum_{j=1}^{n} e^{v_j(x_j)}}. \tag{6}$$

1.3. Models for Discrete Quantity Choices

The discussion of choice models in this section follows the convention of including an outside option that is labeled alternative 0. Individuals in the population make discrete quantity decisions $\mathbf{x} = (x_1, \ldots, x_n)$ where $x_i \in \{0, 1, \ldots, m\}$, and these decisions are observed by the researcher. Decision $\mathbf{x} = \mathbf{0}$ means that the marginal utility of the outside alternative (e.g., alternative 0) dominates the other n alternatives at any positive quantity.

1.3.1. Adaptation of classic discrete choice models

One alternative for modeling such a setting is the MNL model with a choice set that includes all combinations of quantity decisions (Train 2009). For example, if $m = n = 2$, then each individual selects one of $(1 + m)^n = 9$ alternatives corresponding to $\mathbf{X} = \{(0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (1, 2), (2, 0), (2, 1), (2, 2)\}$. The MNL-based model has the advantage of simple probability expressions, e.g.,

$$P\left[\tilde{\mathbf{x}} = \mathbf{x}\right] = \frac{e^{a_{\mathbf{x}}}}{\sum_{\mathbf{y} \in \mathbf{X}} e^{a_{\mathbf{y}}}}$$

where $a_{\mathbf{x}}$ is the parameter that captures the observed characteristics of alternative $\mathbf{x}$. However, the model has the disadvantage of overfitting when estimating the $a_{\mathbf{x}}$ parameters from the data when m is not small, or when observed data cover only a fraction of the quantity-choice set. The utilities associated with alternative $i = 1, \ldots, n$ have $m + 1$ parameters, one for each possible discrete quantity (and $(m + 1)^n - 1$ parameters in total, after normalizing a parameter to value 1). For comparison, MDCEV has up to three parameters (i.e., a_i, b_i, α_i) for each u_i function, reflecting a parsimonious relationship governing marginal utility. This observation motivates exploration of an alternative to MNL-based multiple discrete extreme value model that explicitly accounts for discrete quantity decisions.

1.3.2. Modeling Discrete Diminishing Marginal Utility via Order Statistics

This paper presents one such alternative that is based on order statistics of extreme value random variables. Following Bhatt (2005), I refer to this alternative as a *multiple discrete-discrete extreme value* (MDDEV) model. The model is parsimonious, requiring $n + 1$ parameters and is relatively tractable, admitting closed-form expressions for choice probabilities. In addition, similar to MDCEV, the model recovers MNL probabilities. There are two cases where this occurs. First, if $m = n = 1$, then the choice model is restricted to a binary choice – either $x = 0$ or $x = 1$. In this case, MDDEV choice probabilities reduce to binary logit choice probabilities. Second, the probability of selecting the outside option (i.e., probability $P\left[\tilde{\mathbf{x}} = \mathbf{0}\right]$) matches the MNL probability. This is similar to the result of MDCEV probability (6) matching the MNL probability, e.g., alternative 1 defined as the outside option.

The remainder of this section presents the model. Section 2 presents relevant existing theory. Section 3 develops new theory, leading to choice probability expressions in theorems 1 and 3. Section 4 addresses parameter estimation and Section 5 reflects on the findings. Proofs are located in the appendix.

The utility from selecting $x_i \in \{0, 1, \ldots, m\}$ units of alternative $i \in \{1, \ldots, n\}$ by a randomly selected individual in the population is

$$\tilde{u}_i(x_i) = a_i x_i + \sum_{j=1}^{x_i} \tilde{\varepsilon}_i^{(j)}$$

where

$$\tilde{\varepsilon}_i^{(1)} = \max\{\tilde{\varepsilon}_{i1}, \ldots, \tilde{\varepsilon}_{im}\}$$

$$\tilde{\varepsilon}_i^{(2)} = \max\left\{\{\tilde{\varepsilon}_{i1}, \ldots, \tilde{\varepsilon}_{im}\} \setminus \left\{\tilde{\varepsilon}_i^{(1)}\right\}\right\}$$

$$\tilde{\varepsilon}_i^{(3)} = \max\left\{\{\tilde{\varepsilon}_{i1}, \ldots, \tilde{\varepsilon}_{im}\} \setminus \left\{\tilde{\varepsilon}_i^{(1)}, \tilde{\varepsilon}_i^{(2)}\right\}\right\}$$

$$\vdots$$

$$\tilde{\varepsilon}_i^{(m-1)} = \max\left\{\{\tilde{\varepsilon}_{i1}, \ldots, \tilde{\varepsilon}_{im}\} \setminus \left\{\tilde{\varepsilon}_i^{(1)}, \ldots, \tilde{\varepsilon}_i^{(m-2)}\right\}\right\}$$

$$\tilde{\varepsilon}_i^{(m)} = \min\{\tilde{\varepsilon}_{i1}, \ldots, \tilde{\varepsilon}_{im}\}$$

and $\tilde{\varepsilon}_{ij}$ are iid Gumbel random variables with location parameter $\nu = 0$ and scale parameter β for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Thus, utility is concave in quantity. The marginal utility of alternative i for a randomly selected individual from the population is

$$\Delta\tilde{u}_i(x_i) = a_i + \tilde{\varepsilon}_i^{(x_i)}, \; i \in \{1, \ldots, n\}, \; x_i \in \{1, \ldots, m\}.$$

The marginal utility of the best outside alternative for a randomly selected individual from the population is

$$\Delta\tilde{u}_0 = a_0 + \tilde{\varepsilon}_0,$$

where $\tilde{\varepsilon}_0$ is an independent Gumbel random variable with location parameter $\nu = 0$ and scale parameter β. Figure 1 illustrates expected marginal utility for $m = 25$ and quantity x_i ranges between 1 and 25.

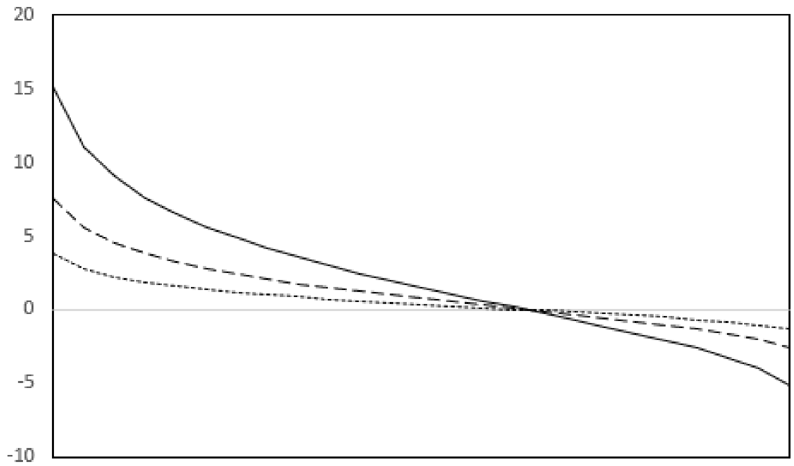

Figure 1. Plot of $E\left[\Delta\tilde{u}_i\left(x_i\right)\right]$ with $m = 25$, $a_i = 0$, and $\beta = 1$ (dotted curve), $\beta = 2$ (dashed curve), $\beta = 3$ (solid curve). (See Lemma 6 for $E\left[\Delta\tilde{u}_i\left(x_i\right)\right]$ expression.)

Individuals select quantities of each alternative to maximize utility. Thus, a randomly selected individual from the population selects x_i or more units of alternative i if and only if

$$\Delta\tilde{u}_i\left(x_i\right) = a_i + \tilde{\varepsilon}_i^{(x_i)} \geq a_0 + \tilde{\varepsilon}_0 = \Delta\tilde{u}_0,\ x_i = 1, \ldots, m$$

and 0 units of alternative i if and only if

$$\Delta\tilde{u}_i\left(1\right) = a_i + \tilde{\varepsilon}_i^{(1)} < a_0 + \tilde{\varepsilon}_0 = \Delta\tilde{u}_0.$$

Without loss of generality, I normalize $a_0 = 0$, e.g., by redefining $a_i = a_i - a_0$ for all i. For example, the fraction of the population that selects x or more units of alternative i is

$$P\left[\tilde{x}_i \geq x\right] = P\left[\Delta\tilde{u}_i\left(x\right) \geq \Delta\tilde{u}_0\right] = P\left[a_i + \tilde{\varepsilon}_i^{(x)} \geq \tilde{\varepsilon}_0\right],\ x \in \{1, \ldots, m\},$$

and

$$P\left[\tilde{x}_i = x\right] = P\left[\Delta\tilde{u}_i\left(x\right) \geq \Delta\tilde{u}_0 > \Delta\tilde{u}_i\left(x+1\right)\right]$$

$$P\left[\tilde{x}_i = 0\right] = P\left[\Delta\tilde{u}_i\left(0\right) < \Delta\tilde{u}_0\right].$$

2. Existing Theory

2.1. Gumbel distribution

The pdf and cdf of a Gumbel random variable $\tilde{z}$ with location parameter ν and scale parameter β are

$$f(z) = \frac{e^{-(z-\nu)/\beta}}{\beta} e^{-e^{-(z-\nu)/\beta}},\ z \in (-\infty, \infty) \tag{7}$$

$$F(z) = e^{-e^{-(z-\nu)/\beta}},\ z \in (-\infty, \infty) \tag{8}$$

with mean and variance

$$E[\tilde{z}] = \nu + \gamma\beta \text{ where } \gamma = -\int_0^\infty e^{-t} \ln t dt \approx 0.577 \text{ is the Euler-Mascheroni constant} \tag{9}$$

$$V[\tilde{z}] = \pi^2\beta^2/6.$$

Lemma 1 (Gumbel 1954). *The Gumbel distribution is closed under maximization. That is, for independent Gumbel random variables* $\tilde{z}_1,...,\tilde{z}_m$ *with scale parameter* β *and location parameters* ν_1, …, ν_m, $\tilde{z}^{(1)} = \max\{\tilde{z}_1,...,\tilde{z}_m\}$ *is a Gumbel random variable with scale parameter* β *and location parameter* $\nu = \beta \ln \sum_{i=1}^{m} e^{\nu_i/\beta}$.

Corollary 1. $\tilde{\varepsilon}_i^{(1)} = \max\{\tilde{\varepsilon}_{i1},...,\tilde{\varepsilon}_{im}\}$ *is a Gumbel random variable with scale parameter* β *and location parameter* $\beta \ln m$.

Lemma 2 (Train 2009, p. 35). *The difference between two independent Gumbel random variables with the same scale parameter is a logistic random variable. That is, for Gumbel* $\tilde{z}_1$ *and* $\tilde{z}_2$ *with scale parameter* β *and location parameters* ν_1 *and* ν_2, $\tilde{z} = \tilde{z}_2 - \tilde{z}_1$ *is logistic with mean* $\nu_2 - \nu_1$, *variance* $\pi^2\beta^2/3$, *and cdf*

$$F_{\tilde{z}}(z) = \frac{1}{1+e^{-(z-(\nu_2-\nu_1))/\beta}}.$$

2.2. Order statistic distribution functions

Let $\tilde{z}_1,...,\tilde{z}_m$ denote independent and identically distributed and continuous random variables with pdf f and cdf F. Let $\tilde{z}^{(1)} = \max\{\tilde{z}_1,...,\tilde{z}_m\}$, $\tilde{z}^{(2)} = \max\{\{\tilde{z}_1,...,\tilde{z}_m\} \setminus \{\tilde{z}^{(1)}\}\}$, $\tilde{z}^{(3)} = \max\{\{\tilde{z}_1,...,\tilde{z}_m\} \setminus \{\tilde{z}^{(1)},\tilde{z}^{(2)}\}\}$, …, $\tilde{z}^{(m)} = \min\{\tilde{z}_1,...,\tilde{z}_m\}$.[1] Then

$$f_{\tilde{z}^{(x)}}(z) = \frac{m!}{(x-1)!(m-x)!} F(z)^{m-x} (1-F(z))^{x-1} f(z) \tag{10}$$

$$F_{\tilde{z}^{(x)}}(z) = \sum_{j=0}^{x-1} \binom{m}{j} F(z)^{m-j} (1-F(z))^{j} \tag{11}$$

$$f_{\tilde{z}^{(1)},...,\tilde{z}^{(x)}}(\mathbf{z}) = \frac{m!}{(m-x)!} F(z_x)^{m-x} \prod_{j=1}^{x} f(z_j),\ z_1 \geq z_2 \geq \ldots \geq z_x,\ x \leq n \tag{12}$$

(see Chapter 2 in David 1981).

[1] This indexing of largest-to-smallest is the reverse of the standard order-statistic convention of smallest-to-largest, but is more convenient for analysis of our model.

3. Theory Development

Define

$A_i = e^{a_i/\beta}$ for i = 0, 1, …, n

(recall that $a_0 = 0$, e.g., $A_0 = 1$). Corollary 2 follows from (7) – (12).

Corollary 2. *For* $x \in \{1, \ldots, m\}$,

$$F_{\Delta\tilde{u}_0}(t) = e^{-e^{-t/\beta}}$$

$$f_{\Delta\tilde{u}_0}(t) = \frac{e^{-t/\beta}}{\beta} e^{-e^{-t/\beta}}$$

$$F_{\Delta\tilde{u}_i(x)}(t) = \sum_{j=0}^{x-1} \binom{m}{j} e^{-(m-j)A_i e^{-t/\beta}} \left(1 - e^{-A_i e^{-t/\beta}}\right)^j$$

$$f_{\Delta\tilde{u}_i(x)}(t) = x \binom{m}{x} \frac{A_i e^{-t/\beta}}{\beta} e^{-(m-x+1)A_i e^{-t/\beta}} \left(1 - e^{-A_i e^{-t/\beta}}\right)^{x-1}$$

$$f_{\Delta\tilde{u}_i(1),\ldots,\Delta\tilde{u}_i(x)}(\mathbf{t}) = \frac{m!}{(m-x)!} \left(\frac{A_i}{\beta}\right)^x e^{-\sum_{j=1}^{x} t_j/\beta} e^{-A_i\left(\sum_{j=1}^{x} e^{-t_j/\beta} + (m-x)e^{-t_x/\beta}\right)}, \; t_1 \geq \ldots \geq t_x.$$

Lemmas 3 – 5 provide the foundation for theorems 1 through 3 that pertain to the probability distribution of random choice vector $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_n)$. Lemma 6 describes the expected value of marginal utility.

Lemma 3. *For any* $i \in \{1, \ldots, n\}$,

$$P[\tilde{x}_i \geq 1] = \frac{mA_i}{1 + mA_i} \tag{13}$$

$$P[\tilde{x}_i = 0] = \frac{1}{1 + mA_i}. \tag{14}$$

Lemma 4. *For any* $i \in \{0, 1, \ldots, n\}$ *and* t,

$$P\left[\Delta\tilde{u}_i(x) \geq t \geq \Delta\tilde{u}_i(x+1)\right] = \binom{m}{x} e^{-(m-x)A_i e^{-t/\beta}} \left(1 - e^{-A_i e^{-t/\beta}}\right)^x, \; x \in \{0, \ldots, m\} \tag{15}$$

$$P\left[\Delta\tilde{u}_i(1) \leq t\right] = P\left[\Delta\tilde{u}_i(0) \geq t \geq \Delta\tilde{u}_i(1)\right] = e^{-mA_i e^{-t/\beta}} \tag{16}$$

$$P\left[\Delta\tilde{u}_i(m) \geq t\right] = P\left[\Delta\tilde{u}_i(m) \geq t \geq \Delta\tilde{u}_i(m+1)\right] = \left(1 - e^{-A_i e^{-t/\beta}}\right)^m \tag{17}$$

$$P\left[\Delta\tilde{u}_i(1) \geq t\right] = 1 - e^{-mA_i e^{-t/\beta}}. \tag{18}$$

Lemma 5. *For a, b* > 0 *and nonnegative integers* $m \geq x$,

$$(m-x)\binom{m}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\left(\frac{b}{a+(m-x+j)b}\right)=\prod_{j=0}^{x}\frac{(m-j)b}{a+(m-j)b}. \tag{19}$$

Lemma 6. *For* $x = 1, \ldots, n$,

$$E\left[\Delta\tilde{u}_i(x)\right]=a_i+\beta x\binom{m}{x}\sum_{j=0}^{x-1}\binom{x-1}{j}(-1)^j\left(\frac{\gamma+\ln(m-x+1+j)}{m-x+1+j}\right) \tag{20}$$

Theorem 1. *Marginal probability functions are*

$$P\left[\tilde{x}_i \geq x\right]=\prod_{j=0}^{x-1}\frac{(m-j)A_i}{1+(m-j)A_i},\ x \in \{1, \ldots, m\} \tag{21}$$

$$P\left[\tilde{x}_i \geq x+1 \mid \tilde{x}_i \geq x\right]=\frac{(m-x)A_i}{1+(m-x)A_i},\ x \in \{1, \ldots, m-1\} \tag{22}$$

$$P\left[\tilde{x}_i = x \mid \tilde{x}_i \geq x\right]=\frac{1}{1+(m-x)A_i},\ x \in \{1, \ldots, m\} \tag{23}$$

$$P\left[\tilde{x}_i = x\right]=\left(\frac{1}{1+(m-x)A_i}\right)\prod_{j=0}^{x-1}\frac{(m-j)A_i}{1+(m-j)A_i},\ x \in \{0, \ldots, m\} \tag{24}$$

$$=\left(\frac{A_i+(m-x)A_i}{1+(m-x)A_i}\right)P\left[\tilde{x}_i = x-1\right],\ x \in \{1, \ldots, m\}. \tag{25}$$

Corollary 3. *If* $A_i = 1$, *then* $P\left[\tilde{x}_i = x\right]=\frac{1}{1+m}$ *for all* $x \in \{0, \ldots, m\}$ *and* $E\left[\tilde{x}_i\right]=\frac{m}{2}$. *If* $A_i < 1$, *then* $P\left[\tilde{x}_i = 0\right] > \ldots > P\left[\tilde{x}_i = m\right]$ *and* $E\left[\tilde{x}_i\right] < \frac{m}{2}$. *If* $A_i > 1$, *then* $P\left[\tilde{x}_i = 0\right] < \ldots < P\left[\tilde{x}_i = m\right]$ *and* $E\left[\tilde{x}_i\right] > \frac{m}{2}$.

Notice that (14) and (24) recover the binary logit choice probabilities when $m = 1$:

$$P\left[\tilde{x}_i = 0\right]=\frac{1}{1+e^{a_i/\beta}}$$

$$P\left[\tilde{x}_i = 1\right]=\frac{e^{a_i/\beta}}{1+e^{a_i/\beta}}.$$

Furthermore, if $n = 1$, so that an individual selects a quantity (possibly zero) up to $m > 1$ of a single alternative, then the probability distribution of random choice $\tilde{x}$ is fully specified by (14) and (24):

$$P\left[\tilde{x} = 0\right]=\frac{1}{1+mA_1}$$

$$P\left[\tilde{x} = 1\right]=\left(\frac{mA_1}{1+(m-1)A_1}\right)P\left[\tilde{x} = 0\right]$$

$$P\left[\tilde{x}=2\right]=\left(\frac{(m-1)A_1}{1+(m-2)A_1}\right)P\left[\tilde{x}=1\right]$$

$$\vdots$$

$$P\left[\tilde{x}=m-1\right]=\left(\frac{2A_1}{1+A_1}\right)P\left[\tilde{x}=m-2\right]$$

$$P\left[\tilde{x}=m\right]=A_1P\left[\tilde{x}=m-1\right]$$

Finally, observe that (22) in Theorem 1 subtly exposes a rather remarkable result on the character of the probability function of conditional Gumbel order statistics, which I state below as a theorem. The result provides a hint of an elegant and curious structure underlying Gumbel order statistics. I come back to this point in Section 5.

Theorem 2. *Let* $\tilde{z}_0,\tilde{z}_1,...,\tilde{z}_m$ *be independent Gumbel random variables with the same scale parameter* β; $\tilde{z}_1,...,\tilde{z}_m$ *have the same location parameter*. *Let* $\tilde{z}^{(x:m)}$ *denote the* x^{th} *largest value in m-dimensional random vector* $(\tilde{z}_1,...,\tilde{z}_m)$. *Then for* $x \in \{1, \ldots, m-1\}$,

$$P\left[\tilde{z}^{(x+1:m)}\geq\tilde{z}_0 \mid \tilde{z}^{(x:m)}\geq\tilde{z}_0\right]=P\left[\tilde{z}^{(1:m-x)}\geq\tilde{z}_0\right].$$

Theorem 3. *The probability mass function of* $\tilde{\mathbf{x}}$ *is*

$$P\left[\tilde{\mathbf{x}}=\mathbf{x}\right]=\left(\prod_{i=1}^{n}\binom{m}{x_i}\right)\sum_{j_1=0}^{x_1}\cdots\sum_{j_n=0}^{x_n}\left(\frac{\binom{x_1}{j_1}\cdots\binom{x_n}{j_n}(-1)^{\left(\sum_{i=1}^{n}j_i\right)}}{1+\sum_{i=1}^{n}(m-x_i+j_i)A_i}\right),\ \mathbf{x}\in\{0,\ldots,m\}^n. \tag{26}$$

or in vector notation,

$$P\left[\tilde{\mathbf{x}}=\mathbf{x}\right]=\left(\prod_{i=1}^{m}\binom{m}{x_i}\right)\vec{C}_2(\mathbf{x})^T\vec{\alpha}_2(\mathbf{x}) \tag{27}$$

where I(•) *is an indicator function returning* 1 *if condition* • *holds and*

$$\vec{C}(\mathbf{x})=\left(\left(\prod_{i=1}^{n}\binom{x_i}{j_i}I(j_i\leq x_i)\right)(-1)^{\sum_{i=1}^{n}j_i}\right)_{\mathbf{j}\in\{0,\ldots,m\}^n}$$

$$\vec{\alpha}(\mathbf{x})=\left(\frac{1}{1+\sum_{i=1}^{n}(m-x_i+j_i)A_i}\right)_{\mathbf{j}\in\{0,\ldots,m\}^n}.$$

The single-choice and null-choice probabilities exhibit a particularly simple structure. For $\mathbf{x}$ = $(x,0,...,0)$ with $x \in \{1, \ldots, m\}$, I apply the identity in Lemma 5, and (26) simplifies to

$$P\left[\tilde{\mathbf{x}}=(x,0,\ldots,0)\right]=\left(\frac{1}{1+m\sum_{i=2}^{n}A_i+(m-x)A_1}\right)\prod_{j=0}^{x-1}\frac{(m-j)A_1}{1+m\sum_{i=2}^{n}A_i+(m-j)A_1}, \tag{28}$$

which follows the marginal probability structure in (24). If $\mathbf{x} = \mathbf{0}$, then (26) simplifies to

$$P\left[\tilde{\mathbf{x}}=\mathbf{0}\right]=\frac{1}{1+m\sum_{i=1}^{n}A_i}, \tag{29}$$

which recovers the MNL null-choice probability.

4. Parameter Estimation

Without loss of generality, the scale parameter is normalized $\beta = 1$, e.g., $A_i = e^{\bar{a}_i/\beta} = e^{a_i}$.[2] For a given value of m, one may estimate vector $\mathbf{A} = (A_1, \ldots, A_n)$ by maximizing the log-likelihood function (i.e., maximum likelihood estimation) of observations $\mathbf{x}^1, \ldots, \mathbf{x}^N$:

$$\begin{aligned}\hat{\mathbf{A}} &= \arg\max_{\mathbf{Y}} \ln\left(\prod_{k=1}^{N} P\left[\tilde{\mathbf{x}}=\mathbf{x}^k \mid \mathbf{A}=\mathbf{Y}\right]\right)\\ &= \arg\max_{\mathbf{Y}}\left(\sum_{k=1}^{N}\ln\left(\left(\prod_{i=1}^{m}\binom{m}{x_i^k}\right)\vec{C}\left(\mathbf{x}^k\right)^T\vec{\alpha}\left(\mathbf{x}^k \mid \mathbf{Y}\right)\right)\right)\end{aligned}$$

where

$$\vec{\alpha}\left(\mathbf{x}^k \mid \mathbf{Y}\right)=\left(\frac{1}{1+\sum_{i=1}^{n}\left(m-x_i^k+j_i\right)Y_i}\right)_{\mathbf{j}\in\{0,\ldots,m\}^n}.$$

An alternative estimation method that provides less theoretical precision (because it ignores some information) but is computationally more efficient is maximum likelihood estimation using the marginal probability functions:

$$\begin{aligned}\hat{\mathbf{A}} &= \arg\max_{\mathbf{Y}} \ln\left(\prod_{i=1}^{n}\prod_{k=1}^{N} P\left[\tilde{x}_i = x_i^k \mid A_i = Y_i\right]\right)\\ &= \arg\max_{\mathbf{Y}} \sum_{i=1}^{n}\sum_{k=1}^{N}\ln\left(P\left[\tilde{x}_i = x_i^k \mid A_i = Y_i\right]\right)\end{aligned}$$

[2] Parameter a_i can be interpreted as the "base" marginal utility of alternative i, which may be modeled as a linear combination of alternative i attributes/characteristics $\bar{\mathbf{a}}_i = \left(\bar{a}_{i1},\ldots,\bar{a}_{iM}\right)$ and a vector $\mathbf{z} = (z_1, \ldots, z_M)$ that reflects characteristics of the population e.g., $a_i = \bar{\mathbf{a}}_i^T\mathbf{z}$.

$$= \arg\max_{\mathbf{Y}} \sum_{i=1}^{n} \sum_{k=1}^{N} \ln\left(\left(\frac{1}{1+\left(m-x_i^k\right)Y_i} \right) \prod_{j=0}^{x_i^k-1} \frac{\left(m-j\right)Y_i}{1+\left(m-j\right)Y_i} \right)$$

$$= \arg\max_{\mathbf{Y}} \sum_{i=1}^{n} \sum_{k=1}^{N} \left(\ln\left(\frac{1}{1+\left(m-x_i^k\right)Y_i} \right) + \sum_{j=0}^{x_i^k-1} \ln\left(\frac{\left(m-j\right)Y_i}{1+\left(m-j\right)Y_i} \right) \right).$$

With this approach, the parameter for each alternative can be estimated independently, i.e., for $i = 1, \ldots, n$,

$$\hat{A}_i = \arg\max_{Y_i} \sum_{k=1}^{N} \left(\ln\left(\frac{1}{1+\left(m-x_i^k\right)Y_i} \right) + \sum_{j=0}^{x_i^k-1} \ln\left(\frac{\left(m-j\right)Y_i}{1+\left(m-j\right)Y_i} \right) \right).$$

I run a small numerical experiment to provide a sense of how the two methods compare in terms of accuracy. I randomly generate individual choices[3] for a model with $m = 3$, $n = 3$, and true parameters $A_1 = 7 \approx e^2$ $A_2 = 1 = e^0$, $A_3 = 0.13 \approx e^{-2}$. I set the sample size to $N = 1{,}000$ and run 30 replications. Table 1 compares the mean absolute percentage error (MAPE) of MLE parameter estimates $\hat{A}_i$ for the two methods. The results illustrate the decrease in precision from ignoring correlation in quantity choices of individuals in a population. In this example, the reduction in accuracy of parameter estimates is relatively small.

		Sample Size = 1000 Choices			
		Method 1		Method 2	
	True Value	Mean MLE Estimate	MAPE (Std Err)	Mean MLE Estimate	MAPE (Std Err)
A_1	7.00	6.942	5.21% (0.61%)	6.874	6.11% (0.69%)
A_2	1.00	1.004	4.21% (0.61%)	1.004	4.23% (0.61%)
A_3	0.13	0.132	5.25% (0.70%)	0.132	5.21% (0.70%)

Table 1. Mean parameter estimate and mean absolute percentage error over 30 replications with sample size $N = 1{,}000$ using the probability mass function (Method 1) and the marginal probability function (Method 2). Standard errors are in parentheses.

5. Discussion

This paper presents a model of individual quantity decisions over a set of alternatives that is consistent with random utility maximization. The model employs order statistics of Gumbel random variables to capture individuals' idiosyncratic and diminishing marginal utility to consumption. The underlying mathematical structure admits closed-form marginal quantity-choice probability functions, and closed-form joint probability functions.

[3] Each of N individuals chooses quantities of each alternative to maximize total utility according to a given realization of $\tilde{\varepsilon}_0$ and $\tilde{\varepsilon}_{ij}$ for $i = 1, \ldots, n$, $j = 1, \ldots, m$.

An important question is whether, or under what settings, the model may capture the essence of real-world behavior . . . whether the model can provide useful predictions or expose new insights for guiding decisions in public or private sector settings. I suspect that the range of meaningful application is narrow, and the reason stems from the property of the probability distribution of Gumbel order statistics identified in Theorem 2. One manifestation of this property is the probability mass function of quantity choice of each alternative is monotonic (see Corollary 3); as A_i increases from 0, the distribution of mass shifts from an extreme right skew (all mass at 0) to uniform at $A_i = 1$, to extreme left skew will all mass at m as A_i approaches infinity. The model restricts the mode of the quantity-choice distribution to be the extreme left, the extreme right, or the entire sample space (in the case of $A_i = 1$), a property that may not fit with reality in many settings.

The monotonic feature of the probability distribution is characteristic of a power-law relation where probability is scale-invariant and proportional to x^{-k}. For example, the continuous analog of the probability distribution in Theorem 1 has density proportional to $(1 + m - x)^k$ where k > (<) 0 if A_i < (>) 1 (see the appendix). A wide variety of phenomena (physical, biological, man-made) exhibit power-law relations.

The result of Theorem 2 identifies a curious phenomenon of Gumbel orders statistics. This phenomenon is perhaps most succinctly illustrated with a few simple probability expressions involving four iid Gumbel random variables, say $\tilde{z}_0$ and $\tilde{\mathbf{z}} = (\tilde{z}_1, \tilde{z}_2, \tilde{z}_3)$:

$$\begin{aligned} P\left[\max\{\tilde{\mathbf{z}}\} \le \tilde{z}_0\right] &= P\left[\max\{\tilde{\mathbf{z}}\} \ge \tilde{z}_0 \ge \max\{\tilde{\mathbf{z}} \setminus \max\{\tilde{\mathbf{z}}\}\}\right] \\ &= P\left[\max\{\tilde{\mathbf{z}} \setminus \max\{\tilde{\mathbf{z}}\}\} \ge \tilde{z}_0 \ge \min\{\tilde{\mathbf{z}}\}\right] \\ &= P\left[\min\{\tilde{\mathbf{z}}\} \ge \tilde{z}_0\right] = 1/4 \end{aligned}$$

$$\frac{P\left[a + \max\{\tilde{\mathbf{z}} \setminus \max\{\tilde{\mathbf{z}}\}\} \ge \tilde{z}_0 \ge a + \min\{\tilde{\mathbf{z}}\}\right]}{P\left[a + \max\{\tilde{\mathbf{z}}\} \ge \tilde{z}_0 \ge a + \max\{\tilde{\mathbf{z}} \setminus \max\{\tilde{\mathbf{z}}\}\}\right]} = \frac{P\left[a + \min\{\tilde{\mathbf{z}}\} \ge \tilde{z}_0\right]}{P\left[a + \max\{\tilde{\mathbf{z}} \setminus \max\{\tilde{\mathbf{z}}\}\} \ge \tilde{z}_0 \ge a + \min\{\tilde{\mathbf{z}}\}\right]} = e^a \; .$$

The expressions show a degree of regularity that hints at an elegant underlying structure that may be illuminating if well understood.

6. References

Bhat CR (2005) A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions. *Transportation Research Part B* 39:679-707.

David HA (1981) *Order Statistics*, 2nd edition, John Wiley & Sons, New York, NY.

Gumbel EJ (1954) *Statistical Theory of Extreme Values and Some Practical Applications*, U.S. Department of Commerce, National Bureau of Standards.

Train KE (2009) *Discrete Choice Methods with Simulation*. Cambridge University Press, New York, NY.

7. Appendix

Proof of Corollary 2. To simplify notation, $f(t) = f_{\Delta\tilde{u}_0}(t)$, $F(t) = F_{\Delta\tilde{u}_0}(t)$, $f_i(t) = f_{a_i+\tilde{\varepsilon}_{ij}}(t)$, and $F_i(t) = F_{a_i+\tilde{\varepsilon}_{ij}}(t)$. From (7) and (8),

$$F(t) = e^{-e^{-t/\beta}} \qquad f(t) = \frac{e^{-t/\beta}}{\beta} e^{-e^{-t/\beta}}$$

$$F_i(t) = e^{-A_i e^{-t/\beta}} \qquad f_i(t) = \frac{A_i e^{-t/\beta}}{\beta} e^{-A_i e^{-t/\beta}}$$

and from (7) – (12),

$$f_{\Delta\tilde{u}_i(x)}(t) = \frac{m!}{(x-1)!(m-x)!} \frac{A_i e^{-t/\beta}}{\beta} e^{-(m-x+1)A_i e^{-t/\beta}} \left(1 - e^{-A_i e^{-t/\beta}}\right)^{x-1}$$

$$F_{\Delta\tilde{u}_i(x)}(t) = \sum_{j=0}^{x-1} \binom{m}{j} e^{-(m-j)A_i e^{-t/\beta}} \left(1 - e^{-A_i e^{-t/\beta}}\right)^{j}$$

$$\begin{aligned}
f_{\Delta\tilde{u}_i(1),\ldots,\Delta\tilde{u}_i(x)}(\mathbf{t}) &= \frac{m!}{(m-x)!} F(t_x)^{m-x} \prod_{j=1}^{x} f(t_j) \\
&= \frac{m!}{(m-x)!} \left(e^{-A_i e^{-t_k/\beta}}\right)^{m-x} \left(\frac{A_i e^{-t_1/\beta}}{\beta} e^{-A_i e^{-t_1/\beta}}\right) \times \cdots \times \left(\frac{A_i e^{-t_x/\beta}}{\beta} e^{-A_i e^{-t_x/\beta}}\right) \\
&= \frac{m!}{(m-x)!} \left(e^{-A_i\left(\sum_{j=1}^{x} e^{-t_j/\beta} + (m-x)e^{-t_x/\beta}\right)} \right) \left(\frac{A_i e^{-t_1/\beta}}{\beta}\right) \times \cdots \times \left(\frac{A_i e^{-t_x/\beta}}{\beta}\right) \\
&= x \binom{m}{x} \left(\frac{A_i}{\beta}\right)^{x} e^{-\sum_{j=1}^{x} t_j/\beta} e^{-A_i\left(\sum_{j=1}^{x} e^{-t_j/\beta} + (m-x)e^{-t_x/\beta}\right)}. \ \square
\end{aligned}$$

Proof of Lemma 3. Recall that $\Delta\tilde{u}_0$ is Gumbel with scale β and location 0. It follows from Lemma 1 that $\Delta\tilde{u}_i(1)$ is an independent Gumbel with scale β and location $a_i + \beta \ln m$. Therefore, from Lemma 2, it follows that $\Delta\tilde{u}_0 - \Delta\tilde{u}_i^{(1)}$ is a logistic random variable with mean $-(a_i + \beta \ln m)$, and thus,

$$\begin{aligned}
P[\tilde{x}_i \geq 1] = P[\Delta\tilde{u}_i(1) \geq \Delta\tilde{u}_0] = P[\Delta\tilde{u}_0 - \Delta\tilde{u}_i(1) \leq 0] &= \frac{1}{1 + e^{-(0-(0-a_i-\beta\ln m))/\beta}} = \frac{1}{1 + e^{-(a_i+\beta\ln m)/\beta}} \\
&= \frac{m e^{a_i/\beta}}{1 + m e^{a_i/\beta}} = \frac{mA_i}{1 + mA_i}
\end{aligned}$$

$$P[\tilde{x}_i = 0] = 1 - P[\tilde{x}_i \geq 1] = \frac{1}{1 + mA_i}. \ \square$$

Proof of Lemma 4. To simplify notation, let $\tilde{\mathbf{u}}_i(x)=\left(\Delta\tilde{u}_i(1),...,\tilde{u}_i(x)\right)$. Then

$$A=P\left[\Delta\tilde{u}_i(x)\ge t\ge\Delta\tilde{u}_i(x+1)\right]=\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-1}}\int_{-\infty}^{t} f_{\tilde{\mathbf{u}}_i(x+1)}\left(t_1,...,t_{x+1}\right)dt_{x+1}...dt_1$$

$$=\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-1}}\int_{-\infty}^{t}\frac{m!}{(m-x-1)!}\left(\frac{A_i}{\beta}\right)^{x+1}e^{-\sum_{j=1}^{x+1}t_j/\beta}e^{-A_i\left(\sum_{j=1}^{x+1}e^{-t_j/\beta}+(m-x)e^{-t_{x+1}/\beta}\right)}dt_{x+1}...dt_1$$

$$=\frac{m!}{(m-x-1)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-1}}\left(\frac{A_i}{\beta}\right)^{x}e^{-\sum_{j=1}^{x}t_j/\beta}e^{-A_i\sum_{j=1}^{x}e^{-t_j/\beta}}\left(\int_{-\infty}^{t}\frac{A_ie^{-t_{x+1}/\beta}}{\beta}e^{-A_i(m-x)e^{-t_{x+1}/\beta}}dt_{x+1}\right)dt_x...dt_1$$

Let $\delta=-A_ie^{-t_{x+1}/\beta}$. Then $d\delta=\frac{A_ie^{-t_{x+1}/\beta}}{\beta}dt$, $t_{x+1}=-\infty\Rightarrow\delta=-\infty$, $t_{x+1}=t\Rightarrow\delta=-A_ie^{-t/\beta}$, and

$$A=\frac{m!}{(m-x-1)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-1}}\left(\frac{A_i}{\beta}\right)^{x}e^{-\sum_{j=1}^{x}t_j/\beta}e^{-A_i\sum_{j=1}^{x}e^{-t_j/\beta}}\left(\int_{-\infty}^{-A_ie^{-t/\beta}}e^{(m-x)\delta}d\delta\right)dt_x...dt_1$$

$$=\frac{m!}{(m-x-1)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-1}}\left(\frac{A_i}{\beta}\right)^{x}e^{-\sum_{j=1}^{x}t_j/\beta}e^{-A_i\sum_{j=1}^{x}e^{-t_j/\beta}}\left(\frac{e^{-(m-x)A_ie^{-t/\beta}}}{m-x}\right)dt_x...dt_1$$

$$=\frac{m!e^{-(m-x)A_ie^{-t/\beta}}}{(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-1}}\left(\frac{A_i}{\beta}\right)^{x}e^{-\sum_{j=1}^{x}t_j/\beta}e^{-A_i\sum_{j=1}^{x}e^{-t_j/\beta}}dt_x...dt_1$$

$$=\frac{m!e^{-(m-x)A_ie^{-t/\beta}}}{(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-2}}\left(\frac{A_i}{\beta}\right)^{x-1}e^{-\sum_{j=1}^{x-1}t_j/\beta}e^{-A_i\sum_{j=1}^{x-1}e^{-t_j/\beta}}\left(\int_t^{t_{x-1}}\frac{A_i}{\beta}e^{-t_x/\beta}e^{-A_ie^{-t_x/\beta}}dt_x\right)dt_{x-1}...dt_1$$

$$=\frac{m!e^{-(m-x)A_ie^{-t/\beta}}}{(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-2}}\left(\frac{A_i}{\beta}\right)^{x-1}e^{-\sum_{j=1}^{x-1}t_j/\beta}e^{-A_i\sum_{j=1}^{x-1}e^{-t_j/\beta}}\left(\int_{-A_ie^{-t/\beta}}^{-A_ie^{-t_{x-1}/\beta}}e^{\delta}d\delta\right)dt_{x-1}...dt_1$$

$$=\frac{m!e^{-(m-x)A_ie^{-t/\beta}}}{(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-2}}\left(\frac{A_i}{\beta}\right)^{x-1}e^{-\sum_{j=1}^{x-1}t_j/\beta}e^{-A_i\sum_{j=1}^{x-1}e^{-t_j/\beta}}\left(e^{-A_ie^{-t_{x-1}/\beta}}-e^{-A_ie^{-t/\beta}}\right)dt_{x-1}...dt_1$$

$$=\frac{m!e^{-(m-x)A_ie^{-t/\beta}}}{(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-3}}\left(\frac{A_i}{\beta}\right)^{x-2}e^{-\sum_{j=1}^{x-2}t_j/\beta}e^{-A_i\sum_{j=1}^{x-2}e^{-t_j/\beta}}\left(\begin{array}{l}\int_t^{t_{x-2}}\frac{A_ie^{-t_{x-1}/\beta}}{\beta}e^{-2A_ie^{-t_{x-1}/\beta}}dt_{x-1}\\ -e^{-A_ie^{-t/\beta}}\int_t^{t_{x-2}}\frac{A_ie^{-t_{x-1}/\beta}}{\beta}e^{-A_ie^{-t_{x-1}/\beta}}dt_{x-1}\end{array}\right)dt_{x-2}...dt_1$$

$$=\frac{m!e^{-(m-x)A_ie^{-t/\beta}}}{(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-3}}\left(\frac{A_i}{\beta}\right)^{x-2}e^{-\sum_{j=1}^{x-2}t_j/\beta}e^{-A_i\sum_{j=1}^{x-2}e^{-t_j/\beta}}\left(\int_{-A_ie^{-t/\beta}}^{-A_ie^{-t_{x-2}/\beta}}e^{2\delta}d\delta-e^{-A_ie^{-t/\beta}}\int_{-A_ie^{-t/\beta}}^{-A_ie^{-t_{x-2}/\beta}}e^{\delta}d\delta\right)dt_{x-2}...dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-3}}\left(\frac{A_i}{\beta}\right)^{x-2} e^{-\sum_{j=1}^{x-2} t_j/\beta}\, e^{-A_i\sum_{j=1}^{x-2} e^{-t_j/\beta}} \left(\begin{array}{l}\dfrac{e^{-2A_i e^{-t_{x-2}/\beta}}-e^{-2A_i e^{-t/\beta}}}{2}\\ -e^{-A_i e^{-t/\beta}}\left(e^{-A_i e^{-t_{x-2}/\beta}}-e^{-A_i e^{-t/\beta}}\right)\end{array}\right)dt_{x-2}\ldots dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-3}}\left(\frac{A_i}{\beta}\right)^{x-2} e^{-\sum_{j=1}^{x-2} t_j/\beta}\, e^{-A_i\sum_{j=1}^{x-2} e^{-t_j/\beta}} \left(\begin{array}{l}\dfrac{e^{-2A_i e^{-t_{x-2}/\beta}}}{2}\\ -e^{-A_i e^{-t/\beta}}e^{-A_i e^{-t_{x-2}/\beta}}+\dfrac{e^{-2A_i e^{-t/\beta}}}{2}\end{array}\right)dt_{x-2}\ldots dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-3}}\left(\frac{A_i}{\beta}\right)^{x-2} e^{-\sum_{j=1}^{x-2} t_j/\beta}\, e^{-A_i\sum_{j=1}^{x-2} e^{-t_j/\beta}}\ \frac{1}{2}\left(e^{-A_i e^{-t_{x-2}/\beta}}-e^{-A_i e^{-t/\beta}}\right)^2 dt_{x-2}\ldots dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{2(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-4}}\left(\frac{A_i}{\beta}\right)^{x-3} e^{-\sum_{j=1}^{x-3} t_j/\beta}\, e^{-A_i\sum_{j=1}^{x-3} e^{-t_j/\beta}} \left(\int_t^{t_{x-3}}\left(\begin{array}{l}\dfrac{A_i e^{-t_{x-2}/\beta}}{\beta}\\ \times e^{-A_i e^{-t_{x-2}/\beta}}\left(\begin{array}{l}e^{-A_i e^{-t_{x-2}/\beta}}\\ -e^{-A_i e^{-t/\beta}}\end{array}\right)^2\end{array}\right)dt_{x-2}\right)dt_{x-3}\ldots dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{2(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-4}}\left(\frac{A_i}{\beta}\right)^{x-3} e^{-\sum_{j=1}^{x-3} t_j/\beta}\, e^{-A_i\sum_{j=1}^{x-3} e^{-t_j/\beta}} \left(\int_{-A_i e^{-t/\beta}}^{-A_i e^{-t_{x-3}/\beta}} e^{\delta}\left(e^{\delta}-e^{-A_i e^{-t/\beta}}\right)^2 d\delta\right)dt_{x-3}\ldots dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{2(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-4}}\left(\frac{A_i}{\beta}\right)^{x-3} e^{-\sum_{j=1}^{x-3} t_j/\beta}\, e^{-A_i\sum_{j=1}^{x-3} e^{-t_j/\beta}} \left(\left.\frac{\left(e^{\delta}-e^{-A_i e^{-t/\beta}}\right)^3}{3}\right|_{-A_i e^{-t/\beta}}^{-A_i e^{-t_{x-3}/\beta}}\right)dt_{x-3}\ldots dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{2(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-4}}\left(\frac{A_i}{\beta}\right)^{x-3} e^{-\sum_{j=1}^{x-3} t_j/\beta}\, e^{-A_i\sum_{j=1}^{x-3} e^{-t_j/\beta}}\ \frac{\left(e^{-A_i e^{-t_{x-3}/\beta}}-e^{-A_i e^{-t/\beta}}\right)^3}{3}dt_{x-3}\ldots dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{3!(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-5}}\left(\frac{A_i}{\beta}\right)^{x-4} e^{-\sum_{j=1}^{x-4} t_j/\beta}\, e^{-A_i\sum_{j=1}^{x-4} e^{-t_j/\beta}} \left(\int_{-A_i e^{-t/\beta}}^{-A_i e^{-t_{x-4}/\beta}} e^{\delta}\left(e^{\delta}-e^{-A_i e^{-t/\beta}}\right)^3 d\delta\right)dt_{x-4}\ldots dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{3!(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-5}}\left(\frac{A_i}{\beta}\right)^{x-4} e^{-\sum_{j=1}^{x-4} t_j/\beta}\, e^{-A_i\sum_{j=1}^{x-4} e^{-t_j/\beta}} \left(\left.\frac{\left(e^{\delta}-e^{-A_i e^{-t/\beta}}\right)^4}{4}\right|_{-A_i e^{-t/\beta}}^{-A_i e^{-t_{x-4}/\beta}}\right)dt_{x-4}\ldots dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{4!(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-5}}\left(\frac{A_i}{\beta}\right)^{x-4} e^{-\sum_{j=1}^{x-4} t_j/\beta}\, e^{-A_i\sum_{j=1}^{x-4} e^{-t_j/\beta}}\left(e^{-A_i e^{-t_{x-4}/\beta}}-e^{-A_i e^{-t/\beta}}\right)^4 dt_{x-4}\ldots dt_1$$

$\vdots$ (repeating the pattern up to t_1, e.g., $x-5=1$, $x-4=2$, $4=x-2$)

$$=\frac{m!e^{-(m-x)A_ie^{-t/\beta}}}{(x-2)!(m-x)!}\int_t^\infty\left(\int_t^{t_1}\left(\frac{A_i}{\beta}\right)^2 e^{-\sum_{j=1}^{2}t_j/\beta}e^{-A_i\sum_{j=1}^{2}e^{-t_j/\beta}}\left(e^{-A_ie^{-t_2/\beta}}-e^{-A_ie^{-t/\beta}}\right)^{x-2}dt_2\right)dt_1$$

$$=\frac{m!e^{-(m-x)A_ie^{-t/\beta}}}{(x-2)!(m-x)!}\int_t^\infty\frac{A_ie^{-t_1/\beta}}{\beta}e^{-A_ie^{-t_1/\beta}}\left(\int_t^{t_1}\frac{A_ie^{-t_2/\beta}}{\beta}e^{-A_ie^{-t_2/\beta}}\left(e^{-A_ie^{-t_2/\beta}}-e^{-A_ie^{-t/\beta}}\right)^{x-2}dt_2\right)dt_1$$

$$=\frac{m!e^{-(m-x)A_ie^{-t/\beta}}}{(x-2)!(m-x)!}\int_t^\infty\frac{A_ie^{-t_1/\beta}}{\beta}e^{-A_ie^{-t_1/\beta}}\left(\int_{-A_ie^{-t/\beta}}^{-A_ie^{-t_1/\beta}}e^{\delta}\left(e^{\delta}-e^{-A_ie^{-t/\beta}}\right)^{x-2}d\delta\right)dt_1$$

$$=\frac{m!e^{-(m-x)A_ie^{-t/\beta}}}{(x-2)!(m-x)!}\int_t^\infty\frac{A_ie^{-t_1/\beta}}{\beta}e^{-A_ie^{-t_1/\beta}}\left(\left.\frac{\left(e^{\delta}-e^{-A_ie^{-t/\beta}}\right)^{x-1}}{x-1}\right|_{-A_ie^{-t/\beta}}^{-A_ie^{-t_1/\beta}}\right)dt_1$$

$$=\frac{m!e^{-(m-x)A_ie^{-t/\beta}}}{(x-1)!(m-x)!}\int_t^\infty\frac{A_ie^{-t_1/\beta}}{\beta}e^{-A_ie^{-t_1/\beta}}\left(e^{-A_ie^{-t_1/\beta}}-e^{-A_ie^{-t/\beta}}\right)^{x-1}dt_1$$

$$=\frac{m!e^{-(m-x)A_ie^{-t/\beta}}}{(x-1)!(m-x)!}\int_{-A_ie^{-t/\beta}}^{0}e^{\delta}\left(e^{\delta}-e^{-A_ie^{-t/\beta}}\right)^{x-1}d\delta$$

$$=\frac{m!e^{-(m-x)A_ie^{-t/\beta}}}{(x-1)!(m-x)!}\left(\left.\frac{\left(e^{\delta}-e^{-A_ie^{-t/\beta}}\right)^{x}}{x}\right|_{-A_ie^{-t/\beta}}^{0}\right)$$

$$=\frac{m!e^{-(m-x)A_ie^{-t/\beta}}}{x!(m-x)!}\left[\left(1-e^{-A_ie^{-t/\beta}}\right)^{x}-\left(e^{-A_ie^{-t/\beta}}-e^{-A_ie^{-t/\beta}}\right)^{x}\right]$$

$$=\binom{m}{x}e^{-(m-x)A_ie^{-t/\beta}}\left(1-e^{-A_ie^{-t/\beta}}\right)^{x},$$

which is (15). Note that

$$P\left[\Delta\tilde{u}_i(1)\le t\right]=\int_{-\infty}^{t}f_{\Delta\tilde{u}_i(1)}(t_1)\,dt_1=\int_{-\infty}^{t}\frac{mA_ie^{-t_1/\beta}}{\beta}e^{-mA_ie^{-t_1/\beta}}dt_1$$

$$=\left(\left.e^{-mA_ie^{-t_1/\beta}}\right|_{-\infty}^{t}\right)=e^{-mA_ie^{-t/\beta}}$$

$$P\left[\Delta\tilde{u}_i(m)\ge t\right]=\int_t^\infty f_{\Delta\tilde{u}_i(m)}(t_1)\,dt_1=\int_t^\infty\frac{mA_ie^{-t_1/\beta}}{\beta}e^{-A_ie^{-t_1/\beta}}\left(1-e^{-A_ie^{-t_1/\beta}}\right)^{m-1}dt_1$$

$$=\left(\left.-\left(1-e^{-A_ie^{-t_1/\beta}}\right)^{m}\right)\right|_t^\infty=\left(1-e^{-A_ie^{-t/\beta}}\right)^{m}$$

and it is apparent that (15) includes (16) and (17) as special cases. Finally,

$$P\left[\Delta\tilde{u}_i\left(1\right)\geq t\right]=1-P\left[\Delta\tilde{u}_i\left(1\right)\leq t\right]=1-e^{-mA_i e^{-t/\beta}}\ . \square$$

Proof of Lemma 5. It is clear that (19) holds at $x = m$. I evaluate over increasing values of x to establish the pattern that generalizes to (19). Let $g(x)$ be the LHS of (19) and $h(x)$ be the RHS of (19). Then for increasing values of $x < m$,

$$g(0)=\frac{mb}{a+mb}=h(0)$$

$$g(1)=m\left(m-1\right)b\left(\frac{1}{a+\left(m-1\right)b}-\frac{1}{a+mb}\right)=\left(\frac{\left(m-1\right)b}{a+\left(m-1\right)b}\right)\left(\frac{mb}{a+mb}\right)=h(1)$$

$$\begin{aligned}
g(2)&=\frac{m!b}{2\left(m-3\right)!}\left(\frac{1}{a+\left(m-2\right)b}-\frac{2}{a+\left(m-1\right)b}+\frac{1}{a+mb}\right)\\
&=\frac{m!b}{2\left(m-3\right)!}\left(\frac{1}{a+\left(m-2\right)b}-\frac{1}{a+\left(m-1\right)b}-\left(\frac{1}{a+\left(m-1\right)b}-\frac{1}{a+mb}\right)\right)\\
&=\frac{m!b}{2\left(m-3\right)!}\left(\frac{b}{\left(a+\left(m-2\right)b\right)\left(a+\left(m-1\right)b\right)}-\frac{b}{\left(a+\left(m-1\right)b\right)\left(a+mb\right)}\right)\\
&=\frac{m!b^2}{2\left(m-3\right)!}\left(\frac{2b}{\left(a+\left(m-2\right)b\right)\left(a+\left(m-1\right)b\right)\left(a+mb\right)}\right)\\
&=\left(\frac{\left(m-2\right)b}{\left(a+\left(m-2\right)b\right)}\right)\left(\frac{\left(m-1\right)b}{a+\left(m-1\right)b}\right)\left(\frac{mb}{a+mb}\right)=h(2)
\end{aligned}$$

$$\begin{aligned}
g(3)&=\frac{m!b}{3!\left(m-4\right)!}\left(\frac{1}{a+\left(m-3\right)b}-\frac{3}{a+\left(m-2\right)b}+\frac{3}{a+\left(m-1\right)b}-\frac{1}{a+mb}\right)\\
&=\frac{m!b}{3!\left(m-4\right)!}\left(\begin{array}{l}\frac{1}{a+\left(m-3\right)b}-\frac{1}{a+\left(m-2\right)b}\\ -2\left(\frac{1}{a+\left(m-2\right)b}-\frac{1}{a+\left(m-1\right)b}\right)+\frac{1}{a+\left(m-1\right)b}-\frac{1}{a+mb}\end{array}\right)\\
&=\frac{m!b^2}{3!\left(m-4\right)!}\left(\begin{array}{l}\frac{1}{\left(a+\left(m-3\right)b\right)\left(a+\left(m-2\right)b\right)}-\frac{1}{\left(a+\left(m-2\right)b\right)\left(a+\left(m-1\right)b\right)}\\ -\left(\frac{1}{\left(a+\left(m-2\right)b\right)\left(a+\left(m-1\right)b\right)}-\frac{1}{\left(a+\left(m-1\right)b\right)\left(a+mb\right)}\right)\end{array}\right)\\
&=\frac{m!b^2}{3!\left(m-4\right)!}\left(\begin{array}{l}\frac{2b}{\left(a+\left(m-3\right)b\right)\left(a+\left(m-2\right)b\right)\left(a+\left(m-1\right)b\right)}\\ -\frac{2b}{\left(a+\left(m-2\right)b\right)\left(a+\left(m-1\right)b\right)\left(a+mb\right)}\end{array}\right)
\end{aligned}$$

$$= \frac{m!b^3}{3(m-4)!}\left(\frac{3b}{(a+(m-3)b)(a+(m-2)b)(a+(m-1)b)(a+mb)}\right)$$

$$= \left(\frac{(m-3)b}{a+(m-3)b}\right)\left(\frac{(m-2)b}{a+(m-2)b}\right)\left(\frac{(m-1)b}{a+(m-1)b}\right)\left(\frac{mb}{a+mb}\right) = h(3)$$

$$\vdots$$

continuing the pattern …

$$\vdots$$

$$g(m-1) = \left(\frac{b}{a+b}\right)\left(\frac{2b}{a+2b}\right)\cdots\left(\frac{(m-1)b}{a+(m-1)b}\right)\left(\frac{mb}{a+mb}\right) = h(m-1). \square$$

Proof of Lemma 6. Recall that $\Delta\tilde{u}_i(x) = a_i + \tilde{\varepsilon}_i^{(x)}$, and thus

$$E\left[\Delta\tilde{u}_i(x)\right] = a_i + E\left[\tilde{\varepsilon}_i^{(x)}\right]$$

where $\tilde{\varepsilon}_i^{(x)}$ is the x^{th} largest value among m independent Gumbel random variables with location parameter $\nu = 0$ and scale parameter β. From Lemma 1, it follows that $E\left[\tilde{\varepsilon}_i^{(1)}\right] = \beta(\ln m + \gamma)$, and thus,

$$E\left[\Delta\tilde{u}_i(1)\right] = a_i + \beta(\ln m + \gamma). \tag{A1}$$

Suppose $x > 1$. For $b > 0$, let $\delta = be^{-t/\beta}$, and note that $d\delta = \frac{-be^{-t/\beta}}{\beta}dt$, $t = \beta\ln\left(\frac{b}{\delta}\right)$, $t = -\infty \Rightarrow \delta = \infty$, $t = \infty \Rightarrow \delta = 0$. Therefore,

$$\int_{-\infty}^{\infty} \frac{tbe^{-t/\beta}}{\beta}e^{-be^{-t/\beta}}dt = -\int_{\infty}^{0}\beta\ln\left(\frac{b}{\delta}\right)e^{-\delta}d\delta = \int_{0}^{\infty}\beta\ln\left(\frac{b}{\delta}\right)e^{-\delta}d\delta = \beta\int_{0}^{\infty}\left[\ln b - \ln\delta\right]e^{-\delta}d\delta$$

$$= \beta\left(-e^{-\delta}\ln b\Big|_0^\infty - \int_0^\infty e^{-\delta}\ln\delta\, d\delta\right) = \beta(\ln b + \gamma). \qquad \text{(see (9))}$$

Note that $f_{\tilde{\varepsilon}_i^{(x)}}(t) = x\binom{m}{x}\frac{e^{-t/\beta}}{\beta}e^{-(m-x+1)e^{-t/\beta}}\left(1-e^{-e^{-t/\beta}}\right)^{x-1}$ (see Corollary 2). Therefore, for $x = 2, \ldots, m$,

$$E\left[\tilde{\varepsilon}_i^{(x)}\right] = \int_{-\infty}^{\infty} tf_{\tilde{\varepsilon}_i^{(x)}}(t)dt = x\binom{m}{x}\int_{-\infty}^{\infty}\frac{te^{-t/\beta}}{\beta}e^{-(m-x+1)e^{-t/\beta}}\left(1-e^{-e^{-t/\beta}}\right)^{x-1}dt.$$

Applying the binomial expansion, $\left(1-e^y\right)^x = \sum_{j=0}^{x}\binom{x}{j}(-1)^j e^{jy}$, $x = 0, 1, 2, 3, \ldots$,

$$E\left[\tilde{\varepsilon}_i^{(x)}\right] = x\binom{m}{x}\int_{-\infty}^{\infty}\frac{te^{-t/\beta}}{\beta}e^{-(m-x+1)e^{-t/\beta}}\sum_{j=0}^{x-1}\binom{x-1}{j}(-1)^j e^{-je^{-t/\beta}}dt$$

$$=x\binom{m}{x}\sum_{j=0}^{x-1}\binom{x-1}{j}\int_{-\infty}^{\infty}\frac{te^{-t/\beta}}{\beta}e^{-(m-x+1+j)e^{-t/\beta}}\left(-1\right)^{j}\,dt$$

$$=\binom{m}{x}\sum_{j=0}^{x-1}\frac{x}{m-x+1+j}\binom{x-1}{j}\int_{-\infty}^{\infty}\frac{t\left(m-x+1+j\right)e^{-t/\beta}}{\beta}e^{-(m-x+1+j)e^{-t/\beta}}\left(-1\right)^{j}\,dt$$

$$=\beta x\binom{m}{x}\sum_{j=0}^{x-1}\binom{x-1}{j}(-1)^{j}\left(\frac{\gamma+\ln\left(m-x+1+j\right)}{m-x+1+j}\right).$$

Thus

$$E\left[\Delta\tilde{u}_i\left(x\right)\right]=a_i+\beta x\binom{m}{x}\sum_{j=0}^{x-1}\binom{x-1}{j}(-1)^{j}\left(\frac{\gamma+\ln\left(m-x+1+j\right)}{m-x+1+j}\right),$$

which, as a consistency check, reduces to (A1) when $x = 1$. □

Proof of Theorem 1. From Corollary 2,

$$A=P\left[\tilde{x}_i\geq x\right]=P\left[\Delta\tilde{u}_i\left(x\right)\geq\Delta\tilde{u}_0\right]=\int_{-\infty}^{\infty}P\left[\Delta\tilde{u}_0\leq t\right]f_{\Delta\tilde{u}_i(x)}\left(t\right)dt$$

$$=\frac{m!}{(x-1)!(m-x)!}\int_{-\infty}^{\infty}e^{-e^{-t/\beta}}\frac{A_ie^{-t/\beta}}{\beta}e^{-(m-x+1)A_ie^{-t/\beta}}\left(1-e^{-A_ie^{-t/\beta}}\right)^{x-1}dt\,.$$

Applying the binomial expansion, $\left(1-e^{y}\right)^{x}=\sum_{j=0}^{x}\binom{x}{j}(-1)^{j}e^{jy}$, $x = 0, 1, 2, 3, \ldots$,

$$A=\frac{m!}{(x-1)!(m-x)!}\int_{-\infty}^{\infty}\frac{A_ie^{-t/\beta}}{\beta}e^{-\left[1+(m-x+1)A_i\right]e^{-t/\beta}}\sum_{j=0}^{x-1}\binom{x-1}{j}(-1)^{j}e^{-jA_ie^{-t/\beta}}dt$$

$$=\frac{m!}{(x-1)!(m-x)!}\sum_{j=0}^{x-1}\int_{-\infty}^{\infty}\frac{A_ie^{-t/\beta}}{\beta}\binom{x-1}{j}(-1)^{j}e^{-\left[1+(m-x+1+j)A_i\right]e^{-t/\beta}}dt\,.$$

Let $\delta=-e^{-t/\beta}$. Then $d\delta=\dfrac{e^{-t/\beta}}{\beta}dt$, $t=-\infty\Rightarrow\delta=-\infty$, $t=\infty\Rightarrow\delta=0$, and

$$A=\frac{m!}{(x-1)!(m-x)!}\sum_{j=0}^{x-1}\int_{-\infty}^{0}\binom{x-1}{j}(-1)^{j}A_ie^{\left[1+(m-x+1+j)A_i\right]\delta}d\delta$$

$$=\left(m-(x-1)\right)\binom{m}{x-1}\sum_{j=0}^{x-1}\binom{x-1}{j}(-1)^{j}\left(\frac{A_i}{1+\left(m-(x-1)+j\right)A_i}\right)$$

and it follows from Lemma 5 that

$$P\left[\tilde{x}_i\geq x\right]=\left(m-(x-1)\right)\binom{m}{x-1}\sum_{j=0}^{x-1}\binom{x-1}{j}(-1)^{j}\left(\frac{A_i}{1+\left(m-(x-1)+j\right)A_i}\right)$$

$$=\prod_{j=0}^{x-1}\frac{(m-j)A_i}{1+(m-j)A_i},$$

which is (21).

Note that $P[\tilde{x}_i \geq 0] = 1$. From Lemma 3,

$$P[\tilde{x}_i \geq 1] = P[\tilde{x}_i \geq 1 \mid \tilde{x}_i \geq 0] = \frac{mA_i}{1+mA_i}.$$

Thus, for $x \in \{1, \ldots, m-1\}$, I rearrange

$$P[\tilde{x}_i \geq x+1] = P[\tilde{x}_i \geq x+1 \mid \tilde{x}_i \geq x] P[\tilde{x}_i \geq x]$$

and apply (21) to obtain

$$P[\tilde{x}_i \geq x+1 \mid \tilde{x}_i \geq x] = \frac{P[\tilde{x}_i \geq x+1]}{P[\tilde{x}_i \geq x]} = \frac{(m-x)A_i}{1+(m-x)A_i},$$

which is (22). Applying the above,

$$P[\tilde{x}_i = x \mid \tilde{x}_i \geq x] = 1 - P[\tilde{x}_i \geq x+1 \mid \tilde{x}_i \geq x] = \frac{1}{1+(m-x)A_i},$$

which is (23). Finally, applying (21) and (23),

$$P[\tilde{x}_i = x] = P[\tilde{x}_i = x \mid \tilde{x}_i \geq x] P[\tilde{x}_i \geq x] = \left(\frac{1}{1+(m-x)A_i}\right) \prod_{j=0}^{x-1} \frac{(m-j)A_i}{1+(m-j)A_i}$$

yields (24), or alternatively (as a consistency check),

$$\begin{aligned} P[\tilde{x}_i = x] &= P[\tilde{x}_i \geq x] - P[\tilde{x}_i \geq x+1] = \prod_{j=0}^{x-1} \frac{(m-j)A_i}{1+(m-j)A_i} - \prod_{j=0}^{x} \frac{(m-j)A_i}{1+(m-j)A_i} \\ &= \left(\frac{1}{1+(m-x)A_i}\right) \prod_{j=0}^{x-1} \frac{(m-j)A_i}{1+(m-j)A_i} \qquad \text{(which is (24))} \\ &= \left(\frac{1}{1+(m-x)A_i}\right)\left(\frac{(m-x+1)A_i}{1+(m-x+1)A_i}\right) \prod_{j=0}^{x-2} \frac{(m-j)A_i}{1+(m-j)A_i} \\ &= \left(\frac{(m-x+1)A_i}{1+(m-x)A_i}\right)\left(\frac{1}{1+(m-x+1)A_i}\right) \prod_{j=0}^{x-2} \frac{(m-j)A_i}{1+(m-j)A_i} \\ &= \left(\frac{(m-x+1)A_i}{1+(m-x)A_i}\right) P[\tilde{x}_i = x-1], \end{aligned}$$

which is (25). □

Proof of Corollary 3. The relationship among probabilities follows directly from (25) and observation that

$$\frac{A_i+(m-x)A_i}{1+(m-x)A_i}\begin{cases}<1, & A_i>1\\ =1, & A_i=1\\ >1, & A_i<1\end{cases}.$$

If $A_i = 1$, then

$$E[\tilde{x}_i]=\frac{1}{1+m}\sum_{j=0}^{m} j=\frac{m}{2} ,$$

and $E[\tilde{x}_i]<\frac{m}{2}$ for $A_i < 1$ and $E[\tilde{x}_i]>\frac{m}{2}$ for $A_i > 1$ follow from the monotonicity of the probability mass function. □

Proof of Theorem 2. To simplify notation I suppress the subscript *i*. Recall that $\Delta\tilde{u}_0=\tilde{\varepsilon}_0$, $\Delta\tilde{u}(x)=a+\tilde{\varepsilon}^{(x)}$, and $\tilde{\varepsilon}^{(x)}$ is the x^{th}-largest value in random vector $(\tilde{\varepsilon}_1,...,\tilde{\varepsilon}_m)$. Random variables $\tilde{\varepsilon}_0,\tilde{\varepsilon}_1,...,\tilde{\varepsilon}_m$ are iid Gumbel random variables with scale parameter β and location parameter normalized to zero. Augment the order statistic superscript to include the dimension of the random vector from which the order statistics are obtained, e.g.,

$$\tilde{\varepsilon}^{(3:m)}=\max\left\{\{\tilde{\varepsilon}_1,...,\tilde{\varepsilon}_m\}\setminus\left\{\tilde{\varepsilon}^{(1:m)},\tilde{\varepsilon}^{(2:m)}\right\}\right\}.$$

For $x \in \{1, \ldots, m-1\}$,

$$P[\tilde{x}\geq x+1 \mid \tilde{x}\geq x]=P\left[\tilde{\varepsilon}_0-\tilde{\varepsilon}^{(x+1:m)}\leq a \mid \tilde{\varepsilon}_0-\tilde{\varepsilon}^{(x:m)}\leq a\right]$$

$$=\frac{(m-x)e^{a/\beta}}{1+(m-x)e^{a/\beta}}. \qquad \text{(see (22))}$$

From Lemma 1, $\tilde{\varepsilon}^{(1:k)}$ is a Gumbel random variable with scale parameter β and location parameter $\nu=\beta\ln k$. From Lemma 2, $\tilde{z}=\tilde{\varepsilon}_0-\tilde{\varepsilon}^{(1:k)}$ is a logistic random variable with cdf

$$F_{\tilde{z}}(z)=\frac{1}{1+e^{-(z-(0-\beta\ln k))/\beta}}=\frac{1}{1+e^{-z/\beta+\ln k}}=\frac{ke^{z/\beta}}{1+ke^{z/\beta}}.$$

Thus,

$$P\left[\tilde{\varepsilon}_0-\tilde{\varepsilon}^{(x+1:m)}\leq a \mid \tilde{\varepsilon}_0-\tilde{\varepsilon}^{(x:m)}\leq a\right]=P\left[\tilde{\varepsilon}_0-\tilde{\varepsilon}^{(1:m-x)}\leq a\right]. \text{ □}$$

Proof of Theorem 3. From Corollary 2 and Lemma 4,

$$A=P[\tilde{\mathbf{x}}=\mathbf{x}]=\int_{-\infty}^{\infty} f_{\Delta\tilde{u}_0}(t)\left(\prod_{i=1}^{n}P\left[\Delta\tilde{u}_i(x_i)\geq t\geq\Delta\tilde{u}_i(x_i+1)\right]\right)dt$$

$$=\int_{-\infty}^{\infty}\frac{e^{-t/\beta}}{\beta}e^{-e^{-t/\beta}}\left(\prod_{i=1}^{n}\binom{m}{x_i}e^{-(m-x_i)A_ie^{-t/\beta}}\left(1-e^{-A_ie^{-t/\beta}}\right)^{x_i}\right)dt$$

$$=\left(\prod_{i=1}^{n}\binom{m}{x_i}\right)\int_{-\infty}^{\infty}\frac{e^{-t/\beta}}{\beta}\left(e^{-\left(1+\sum_{i=1}^{n}(m-x_i)A_i\right)e^{-t/\beta}}\right)\left(\prod_{i=1}^{n}\left(1-e^{-A_i e^{-t/\beta}}\right)^{x_i}\right)dt\,.$$

Applying the binomial expansion,

$$\left(1-e^{y}\right)^{x}=\sum_{j=0}^{x}\binom{x}{j}(-1)^{j}e^{jy}\text{ , } x = 0, 1, 2, 3, \ldots,$$

$$A=\left(\prod_{i=1}^{n}\binom{m}{x_i}\right)\int_{-\infty}^{\infty}\frac{e^{-t/\beta}}{\beta}\left(e^{-\left(1+\sum_{i=1}^{n}(m-x_i)A_i\right)e^{-t/\beta}}\right)\left(\sum_{j_1=0}^{x_1}\cdots\sum_{j_n=0}^{x_n}\binom{x_1}{j_1}\cdots\binom{x_n}{j_{kn}}(-1)^{\left(\sum_{i=1}^{n}j_i\right)}e^{-\left(\sum_{i=1}^{n}j_iA_i\right)e^{-t/\beta}}\right)dt$$

$$=\left(\prod_{i=1}^{n}\binom{m}{x_i}\right)\left(\sum_{j_1=0}^{x_1}\cdots\sum_{j_n=0}^{x_n}\int_{-\infty}^{\infty}\binom{x_1}{j_1}\cdots\binom{x_n}{j_n}(-1)^{\left(\sum_{i=1}^{n}j_i\right)}\frac{e^{-t/\beta}}{\beta}\left(e^{-\left(1+\sum_{i=1}^{n}(m-x_i)A_i\right)e^{-t/\beta}}\right)e^{-\left(\sum_{i=1}^{n}j_iA_i\right)e^{-t/\beta}}dt\right)$$

$$=\left(\prod_{i=1}^{n}\binom{m}{x_i}\right)\left(\sum_{j_1=0}^{x_1}\cdots\sum_{j_n=0}^{x_n}\int_{-\infty}^{\infty}\binom{x_1}{j_1}\cdots\binom{x_n}{j_n}(-1)^{\left(\sum_{i=1}^{n}j_i\right)}\frac{e^{-t/\beta}}{\beta}\left(e^{-\left(1+\sum_{i=1}^{n}(m-x_i+j_i)A_i\right)e^{-t/\beta}}\right)dt\right).$$

Let $\delta=-e^{-t/\beta}$. Then $d\delta=\dfrac{e^{-t/\beta}}{\beta}dt$, $t=-\infty\Rightarrow\delta=-\infty$, $t=\infty\Rightarrow\delta=0$, and

$$A=P\left[\tilde{\mathbf{x}}=\mathbf{x}\right]=\left(\prod_{i=1}^{n}\binom{m}{x_i}\right)\left(\sum_{j_1=0}^{x_1}\cdots\sum_{j_n=0}^{x_k}\int_{-\infty}^{0}\binom{x_1}{j_1}\cdots\binom{x_n}{j_n}(-1)^{\left(\sum_{i=1}^{n}j_i\right)}\left(e^{\left(1+\sum_{i=1}^{n}(m-x_i+j_i)A_i\right)\delta}\right)d\delta\right)$$

$$=\left(\prod_{i=1}^{n}\binom{m}{x_i}\right)\sum_{j_1=0}^{x_1}\cdots\sum_{j_n=0}^{x_n}\left(\frac{\binom{x_1}{j_1}\cdots\binom{x_n}{j_n}(-1)^{\left(\sum_{i=1}^{n}j_i\right)}}{1+\sum_{i=1}^{n}(m-x_i+j_i)A_i}\right).$$

As a consistency check, recall that $\Delta\tilde{u}_0,\Delta\tilde{u}_1(1),\ldots,\Delta\tilde{u}_n(1)$ are independent Gumbel random variables with scale β and location parameters 0 and $a_i+\beta\ln m$ for $i = 1, \ldots, n$. Thus, it follows from lemmas 1 and 2 that

$$P\left[\tilde{\mathbf{x}}=\mathbf{0}\right]=P\left[\Delta\tilde{u}_0\geq\max_i\left\{\Delta\tilde{u}_i(1)\right\}\right]=\frac{1}{1+e^{-\left(0-\left(\sum_{i=1}^{n}(a_i+\beta\ln m)-0\right)\right)/\beta}}=\frac{1}{1+m\sum_{i=1}^{n}A_i},$$

which aligns with the general expression above. □

Continuous analog of the discrete distribution in Theorem 1.[4] Define $g(x) = P\left[\tilde{x} = x\right]$, $x = 0, \ldots, m$.

From Theorem 1 (with the subscript for alternative *i* suppressed),

$$g(0) = \frac{1}{1+mA}$$

$$g(x) = \left(\frac{1+m-x}{1/A+m-x}\right) g\left(x-1\right),\ x = 1, \ldots, m.$$

Thus,

$$\frac{\Delta g\left(x\right)}{g\left(x\right)} = \frac{g\left(x\right)-g\left(x-1\right)}{g\left(x\right)} = \frac{1+m-x-\left(1/A+m-x\right)}{1+m-x} = \frac{1-1/A}{1+m-x}.$$

The continuous analog, $h(x)$, satisfies

$$\frac{h'\left(x\right)}{h\left(x\right)} = \frac{1-1/A}{1+m-x}.$$

For $A \neq 1$, the solution to the differential equation is

$$h(x) = c\left(1+m-x\right)^{1/A-1}$$

where c is the positive constant that assures $h(x)$ integrates to one over its support.

[4] This continuous analog was derived by Harish Guda, W. P. Carey School of Business, Arizona State University.