Choice-Theoretic Foundations of the Divisive Normalization Model: Online Appendix

November 9, 2018

In this online appendix to "Choice-Theoretic Foundations of the Divisive Normalization Model", we show an equivalence result between: (1) the normalization model with a more general functional form for the divisive factor; (2) the information-processing model without the MCC; and (3) Axioms 1 and 2. This equivalence was referred to several times in the text of the main paper.

Definition 1. A random choice rule ρ has a **generalized divisive normalization representation (GDNR)** if there exists $v : X \to \mathbb{R}_{++}$ and $F : \mathcal{A} \to \mathbb{R}_{++}$ such that for any $A \in \mathcal{A}$ and $x \in A$

$$\rho\left(x, A\right) = \Pr\left(x \in \arg\max_{y \in A} \frac{v\left(y\right)}{F\left(A\right)} + \varepsilon_y\right),$$

where ε_y is distributed i.i.d. Gumbel (0,1).

Theorem 1. *For any random choice rule ρ the following are equivalent:*

1. *ρ has a GDNR,*
2. *ρ has an information-processing representation,*
3. *ρ obeys Axioms 1 and 2.*

We will show that all three parts of Theorem 1 are equivalent to the following statement: there exists $v : X \to \mathbb{R}_{++}$ and $F : \mathcal{A} \to \mathbb{R}_{++}$ such that

$$\rho\left(x, A\right) = \frac{\exp\left(\frac{v(x)}{F(A)}\right)}{\sum_{y \in A} \exp\left(\frac{v(y)}{F(A)}\right)} \text{ for all } A \in \mathcal{A} \text{ and } x \in A. \tag{1}$$

As a shorthand, we will refer to this statement as Equation (1) holding. The proof of the equivalence between Equation (1) and the GDNR follows closely the proof in Appendix A.1

from the main paper, with $F(A)$ replacing $\gamma/(\sigma + v(A))$. We now prove the equivalence to the other two parts.

Equivalence with the Information-Processing Model

First suppose ρ obeys Equation (1) using (v, F). Define an information-processing model using the same v and setting

$$C_A(c) = F(A)\, c$$

for all $c \in \mathbb{R}$ and $A \in \mathcal{A}$. Fix an $A \in \mathcal{A}$. On A, the information-processing model defines a maximization problem with a continuous objective function and a compact constraint set. Hence, there exists a solution $p \in \Delta A$. For each $A \in \mathcal{A}$, p must obey the Karush-Kuhn-Tucker condition that there exists a λ such that for each $x \in A$

$$v(x) - C'_A(\Delta H(p))(\ln(p(x)) + 1) + \lambda + \mu_x = 0, \tag{2}$$

for some μ_x, with the complimentary slackness condition that $\mu_x > 0 \Rightarrow p(x) = 0$. Applying our definition of C_A gives

$$v(x) - F(A)(\ln(p(x)) + 1) + \lambda + \mu_x = 0.$$

We know that $p(x) > 0$ since otherwise the left-hand side is infinite and this equation can never hold. Thus $\mu_x = 0$ by complimentary slackness. It then follows that for any $x, y \in A$

$$\frac{p(x)}{p(y)} = \exp\left(\frac{v(x) - v(y)}{F(A)}\right).$$

And using the fact that probabilities sum to 1, we can derive

$$p(x) = \frac{\exp\left(\frac{v(x)}{F(A)}\right)}{\sum_{y \in A} \exp\left(\frac{v(y)}{F(A)}\right)}$$

for all $x \in A$. It follows that $p = \rho(\cdot, A)$. Repeating this argument for all choice sets, it follows that ρ has an information-processing representation.

Now suppose ρ has an information-processing representation $\left(v, \{C_A\}_{A \in \mathcal{A}}\right)$. For each $A \in \mathcal{A}$, define

$$F(A) = C'_A(\Delta H(\rho(\cdot, A))).$$

We can then write the Karush-Kuhn-Tucker conditions that ρ must obey as

$$v(x) - F(A)(\ln(\rho(x, A)) + 1) + \lambda + \mu_x = 0.$$

We know that $\mu_x = 0$ since $\rho(x, A) > 0$ by assumption. We can now use the same steps as in the proof of the other direction to establish that ρ obeys Equation (1).

Equivalence with Axioms 1 and 2

First, suppose ρ obeys Equation (1) using (v, F). Let $x, y \in A \cap B$. Then $\rho(x, A) \geq \rho(y, A)$ if and only if $v(x) \geq v(y)$ if and only if $\rho(x, B) \geq \rho(y, B)$, which establishes Axiom 1. Next, let (x, y) be distinguishable in A. By definition,

$$R_{xy}(A) := \left(\ln \frac{\rho(x, A)}{\rho(y, A)}\right)^{-1} \ln \frac{\rho(x, X)}{\rho(y, X)}.$$

Applying Equation (1) to the right-hand side gives us that

$$R_{xy}(A) = \frac{F(X)}{F(A)},$$

which does not depend on the choice of (x, y), and Axiom 2 follows.

Now suppose ρ obeys Axioms 1 and 2. We want to show that Equation (1) holds. By Axiom 1, if (x, y) is distinguishable in one set, then (x, y) is distinguishable in all sets that contain this pair. Hence, we will simply say (x, y) is distinguishable to indicate that $\rho(x, A) \neq \rho(y, A)$ whenever $x, y \in A$. By Axiom 2, for any $A \in \mathcal{A}$, we can set $R(A) = R_{xy}(A)$ for all distinguishable (x, y) in A. If A does not contain any distinguishable pairs, set $R(A) = 1$. By Axiom 1, the two natural log terms in the definition of $R_{xy}(A)$ have the same sign. Moreover, if (x, y) is distinguishable then neither natural log term is zero. Thus, $R(A) > 0$ holds for all $A \in \mathcal{A}$.

Define

$$\alpha := 1 - \min_{x \in X} \ln \rho(x, X),$$

and define $v : X \to \mathbb{R}_{++}$ as

$$v(x) := \alpha + \ln \rho(x, X).$$

The construction of α ensures that $v(x) > 0$ for all $x \in X$.

Next define $F : \mathcal{A} \to \mathbb{R}_{++}$ as

$$F(A) = R(A).$$

For all $A \in \mathcal{A}$, $F(A) > 0$ since $R(A) > 0$.

Now choose any $A \in \mathcal{A}$. It suffices to show that for each $x, y \in A$

$$\frac{\rho(x, A)}{\rho(y, A)} = \exp\left(\frac{v(x) - v(y)}{F(A)}\right). \tag{3}$$

Showing this is enough because we can derive Equation (1) using the fact that probabilities sum to 1. Fix a pair (x, y). If (x, y) is not distinguishable, then we know $v(x) = v(y)$. In this case, Equation (3) holds because both sides are equal to 1. Now suppose (x, y) is distinguishable. We can rewrite Equation (3) as

$$F(A) \ln \frac{\rho(x, A)}{\rho(y, A)} = v(x) - v(y).$$

Since A contains a distinguishable pair and $F(A) = R(A)$, this is equivalent to

$$\ln \frac{\rho(x, X)}{\rho(y, X)} = v(x) - v(y),$$

which holds by the definition of $v(x)$, so we are done.