Mathematical details for the uniqueness of the conditional ordinal stereotype model

Greg Ridgeway
Department of Criminology
Department of Statistics and Data Science
University of Pennsylvania

January 5, 2026

This document provides extended mathematical background for the conditional ordinal stereotype model described in Ridgeway (nd). It shows that Anderson's ordinal stereotype model (Anderson, 1984) is the only ordinal model for which the conditional likelihood for officer fixed effects is invariant to environmental features.

From a set of use-of-force incidents we wish to model the probability that officer i uses force type y in an environment characterized by $\mathbf{z}$

$$P(Y = y|\mathbf{z}, \lambda_i) = f(h_y(\mathbf{z}) + s_y\lambda_i) \tag{1}$$

where $y \in \{0, 1, 2, \ldots, J\}$ representing increasing levels of force type severity where 0 represents no force (or a witness officer) and J represents lethal or potentially lethal force. The h_y are functions of the environmental features and λ is the officer fixed effect, the individual officer's propensity to escalate force, with s_y as a coefficient.

For this application, the relative risk of using a more serious force type must increase with λ. If $y_2 > y_1$ then

$$\frac{P(Y = y_2|\mathbf{z}, \lambda + \epsilon)}{P(Y = y_1|\mathbf{z}, \lambda + \epsilon)} > \frac{P(Y = y_2|\mathbf{z}, \lambda)}{P(Y = y_1|\mathbf{z}, \lambda)} \tag{2}$$

Primary challenges in estimating this model unconditionally include

- $\mathbf{z}$ is extremely complex and hard to measure. At a minimum it is difficult to enumerate and measure a broadly agreed upon list of environmental features
- h is likely highly complex
- No generally available data from times and places where $y = 0$ (no force used, police engaged in general patrol and investigative activities)
- No obvious choice for an ordinal regression model to use for f

With a random sample of times and places where an officer is (including both use-of-force incidents and moments in which no officer used force), the unconditional likelihood would have the form

$$L(\boldsymbol{\lambda}, \mathbf{h}, \mathbf{s}) = \prod_{\ell=1}^{n} P\big(Y = y_\ell|\mathbf{z}_\ell, \lambda_{\mathrm{id}(\ell)}\big) = \prod_{\ell=1}^{n} f\big(h_{y_\ell}(\mathbf{z}_\ell) + s_{y_\ell}\lambda_{\mathrm{id}(\ell)}\big) \tag{3}$$

where $\mathrm{id}(\ell)$ gives the index of which officer used force type y_ℓ. Because officers often share environmental features, $\mathbf{z}$, conditional likelihood seems a viable direction. Match officers that share a sampled time and place

$$CL(\boldsymbol{\lambda}, \mathbf{h}, \mathbf{s}) = \prod_{\ell=1}^{n} P\big(y_{1\ell}, \ldots, y_{m_\ell\ell}|\mathbf{z}_\ell, \lambda_{\mathrm{id}(1,\ell)}, \ldots, \lambda_{\mathrm{id}(m_\ell,\ell)}, \mathbf{T}_\ell\big) \tag{4}$$

where $\mathbf{T}_\ell$ is a sufficient statistic for h.

Within a matched incident, the effect of the shared environment enters the likelihood in a manner that is invariant to permutations of officer indices. Consequently, any statistic $\mathbf{T}_\ell$ used to construct a conditional likelihood that removes this shared component must itself be invariant to such permutations. The appropriate conditioning object is, therefore, a maximal invariant under permutations of the officer labels. Theorem 1 shows that the force-type count vector is a maximal invariant.

Theorem 1: Force type counts are the maximal invariant

Let

$$\mathbf{y} = [y_1, \ldots, y_m] \in \{0, 1, \ldots, J\}^m \tag{5}$$

denote the vector of ordinal outcomes for a matched incident with m officers. Define the count vector

$$\mathbf{k}(\mathbf{y}) = [k_0(\mathbf{y}), \ldots, k_J(\mathbf{y})], \qquad k_j(\mathbf{y}) = \sum_{i=1}^{m} \mathbb{I}(y_i = j), \quad j = 0, 1, \ldots, J \tag{6}$$

Let S_m act on $\{0, 1, \ldots, J\}^m$ by permuting indices,

$$(\pi \cdot \mathbf{y})_i = y_{\pi(i)}, \qquad \pi \in S_m. \tag{7}$$

For any $\mathbf{y}, \mathbf{y}' \in \{0, 1, \ldots, J\}^m$,

$$\mathbf{k}(\mathbf{y}) = \mathbf{k}(\mathbf{y}') \iff \exists\, \pi \in S_m \text{ such that } \mathbf{y}' = \pi \cdot \mathbf{y}. \tag{8}$$

Hence $\mathbf{k}(\mathbf{y})$ is a maximal invariant under the action of S_m.

Proof. Proving Theorem 1 requires showing 1) $\mathbf{k}(\mathbf{y})$ is invariant under permutations and 2) $\mathbf{k}(\mathbf{y}) = \mathbf{k}(\mathbf{y}')$ implies there is a $\pi \in S_m$ such that $\mathbf{y}' = \pi \cdot \mathbf{y}$.

To show that $\mathbf{k}(\mathbf{y})$ is invariant under permutations, choose $\pi \in S_m$ and $j \in \{0, 1, \ldots, J\}$. Then

$$\begin{aligned} k_j(\pi \cdot \mathbf{y}) &= \sum_{i=1}^{m} \mathbb{I}\big((\pi \cdot \mathbf{y})_i = j\big) \\ &= \sum_{i=1}^{m} \mathbb{I}\big(y_{\pi(i)} = j\big) \\ &= \sum_{i^*=1}^{m} \mathbb{I}\big(y_{i^*} = j\big) \\ &= k_j(\mathbf{y}) \end{aligned} \tag{9}$$

where the third equality follows since π is a permutation of $\{1, \ldots, m\}$. Since (9) holds for every j, we have $\mathbf{k}(\pi \cdot \mathbf{y}) = \mathbf{k}(\mathbf{y})$.

To prove maximality, assume $\mathbf{k}(\mathbf{y}) = \mathbf{k}(\mathbf{y}')$. For each $j \in \{0, 1, \ldots, J\}$ define the sets containing the indices of which officer used force type j

$$I_j = \{i \in \{1, \ldots, m\} : y_i = j\}, \qquad I'_j = \{i \in \{1, \ldots, m\} : y'_i = j\}. \tag{10}$$

Then $|I_j| = k_j(\mathbf{y}) = k_j(\mathbf{y}') = |I'_j|$. Since I_j and I'_j have the same number of elements, there exists a bijection $\varphi_j : I'_j \to I_j$ for each j. Since $\{I_j\}_{j=0}^{J}$ partitions $\{1, \ldots, m\}$, we may define a permutation $\varphi \in S_m$ by $\varphi(i) = \varphi_j(i)$ whenever $i \in I'_j$. If $i \in I'_j$ then $\varphi(i) \in I_j$, so $\mathbf{y}'_i = y_{\varphi(i)}$ for all i. Therefore $\mathbf{y}' = \varphi \cdot \mathbf{y}$, establishing the implication

$$\mathbf{k}(\mathbf{y}) = \mathbf{k}(\mathbf{y}') \implies \exists\, \pi \in S_m \text{ such that } \mathbf{y}' = \pi \cdot \mathbf{y}. \tag{11}$$

The reverse implication follows from invariance already shown. If $\mathbf{y}' = \pi \cdot \mathbf{y}$, then $\mathbf{k}(\mathbf{y}') = \mathbf{k}(\pi \cdot \mathbf{y}) = \mathbf{k}(\mathbf{y})$. This proves the equivalence in the theorem and hence that $\mathbf{k}(\mathbf{y})$ is a maximal invariant. $\square$

Corollary 2: Any permutation-invariant statistic $\mathbf{T}$ is a function of $\mathbf{k}$

Let $\mathbf{T}(y)$ be any statistic satisfying $\mathbf{T}(\pi \cdot y) = \mathbf{T}(y)$ for all $\pi \in S_m$. Then there exists a function ψ such that

$$\mathbf{T}(y) = \psi\big(\mathbf{k}(y)\big). \tag{12}$$

Proof. Fix y and consider the set $\{y' : \mathbf{k}(y') = \mathbf{k}(y)\}$. By Theorem 1, this set is exactly the orbit of y under the action of S_m. Since $\mathbf{T}(\pi \cdot y) = \mathbf{T}(y)$, it is constant on the orbit, hence constant on each level set of $\mathbf{k}$. For any feasible count vector $\mathbf{k}$, define $\psi(\mathbf{k}) = \mathbf{T}(\mathbf{y})$ for any $\mathbf{y}$ such that $\mathbf{k}(\mathbf{y}) = \mathbf{k}$. This is well-defined since $\mathbf{T}$ is constant on the level set. $\square$

The relevance of Theorem 1 for inference is that a conditional likelihood constructed to remove the shared incident-level component must condition on a statistic that is invariant to permutations of officer indices. According to Corollary 2, any such statistic is necessarily a function of the force-type count vector $\mathbf{k}$. Conditioning on $\mathbf{k}$ is the canonical way to form a conditional likelihood that respects the permutation invariance induced by the matched-incident design.

Proposition 3: Conditional likelihood formulation

If

1. the distribution of Y_i depends on $\mathbf{z}$ and officer i in an additive model of the form

$$P(Y_i = y|\mathbf{z}, \lambda_i) = f(y, h_y(\mathbf{z}) + s_y\lambda_i) \tag{13}$$

2. $Y_1, \ldots, Y_m$, outcomes for m matched officers with a shared value for $\mathbf{z}$, are conditionally independent

3. $\mathcal{K}(\mathbf{k}) = \{y^* \in \{0, 1, \ldots, J\}^m : \mathbf{k}(y^*) = \mathbf{k}\}$ (equivalently, the orbit of $\mathbf{y}$ under permutations of the officer indices)

then

$$CL(\mathbf{h}, \mathbf{s}, \boldsymbol{\lambda} \mid \mathbf{k}) = \frac{\prod_{i=1}^m f\big(y_i, h_{y_i}(\mathbf{z}) + s_{y_i}\lambda_i\big)}{\sum_{\mathbf{y}^* \in \mathcal{K}} \prod_{i=1}^m f\big(y_i^*, h_{y_i^*}(\mathbf{z}) + s_{y_i^*}\lambda_i\big)} \tag{14}$$

Proof.

$$\begin{aligned}
CL(\mathbf{h}, \mathbf{s}, \boldsymbol{\lambda} \mid \mathbf{y}, \mathbf{k}, \mathbf{z}) &= P(Y_1 = y_1, \ldots, Y_m = y_m \mid \mathbf{k}, \mathbf{z}, \mathbf{h}, \mathbf{s}, \boldsymbol{\lambda}) & (15)\\
&= \frac{P(\mathbf{k} \mid y_1, \ldots, y_m, \mathbf{z}, \mathbf{h}, \mathbf{s}, \boldsymbol{\lambda}) P(y_1, \ldots, y_m \mid \mathbf{z}, \mathbf{h}, \mathbf{s}, \boldsymbol{\lambda})}{\sum_{\mathbf{y}^* \in \mathcal{K}(\mathbf{k})} P(\mathbf{k} \mid y_1^*, \ldots, y_m^*, \mathbf{z}, \mathbf{h}, \mathbf{s}, \boldsymbol{\lambda}) P(y_1^*, \ldots, y_m^* \mid \mathbf{z}, \mathbf{h}, \mathbf{s}, \boldsymbol{\lambda})} & (16)\\
&= \frac{1 \cdot P(y_1 \mid \mathbf{z}, \mathbf{h}, \mathbf{s}, \lambda_1) \cdots P(y_m \mid \mathbf{z}, \mathbf{h}, \mathbf{s}, \lambda_m)}{\sum_{\mathbf{y}^* \in \mathcal{K}(\mathbf{k})} 1 \cdot P(y_1^* \mid \mathbf{z}, \mathbf{h}, \mathbf{s}, \lambda_1) \cdots P(y_m^* \mid \mathbf{z}, \mathbf{h}, \mathbf{s}, \lambda_m)} & (17)\\
&= \frac{\prod_{i=1}^m f(y_i, h_{y_i}(\mathbf{z}) + s_{y_i}\lambda_i)}{\sum_{\mathbf{y}^* \in \mathcal{K}(\mathbf{k})} \prod_{i=1}^m f(y_i^*, h_{y_i^*}(\mathbf{z}) + s_{y_i^*}\lambda_i)} & (18)
\end{aligned}$$

where (16) follows from Bayes Theorem, the $P(\mathbf{k}|\cdot)$ reduces to 1 in (17) because the values of $\mathbf{y}$ determine $\mathbf{k}$, and the conditional independence assumption allows the factoring in (17). $\square$

The remaining step is characterizing the class of functions f that remove dependence on $h_y(\mathbf{z})$ from (18).

Theorem 4: Multiplicative separability is necessary and sufficient for h-invariant conditional likelihood

Assume $f(y,t) > 0$ is continuously differentiable in t. Fix $s_0 = 0$ and $s_1 = 1$ for identifiability. For $m \geq 2$ and any choice of $\boldsymbol{\lambda} \in \mathbb{R}^m$, the conditional likelihood

$$\frac{\prod_{i=1}^{m} f(y_i, h_{y_i}(\mathbf{z}) + s_{y_i}\lambda_i)}{\sum_{\mathbf{y}^* \in \mathcal{K}} \prod_{i=1}^{m} f(y_i^*, h_{y_i^*}(\mathbf{z}) + s_{y_i^*}\lambda_i)} \tag{19}$$

is independent of the environmental components, $h_y(\mathbf{z})$, for all $\mathbf{z}$ if and only if

$$f(y, h_y(\mathbf{z}) + s_y\lambda) = g(y, h_y(\mathbf{z}))G(y, s_y\lambda) \tag{20}$$

for $1 \leq y \leq J$ and $g, G : (\mathbb{N}_0, \mathbb{R}) \to \mathbb{R}_+$, so that the conditional likelihood simplifies to

$$\frac{\prod_{i=1}^{m} G(y_i, s_{y_i}\lambda_i)}{\sum_{\mathbf{y}^* \in \mathcal{K}} \prod_{i=1}^{m} G(y_i^*, s_{y_i^*}\lambda_i)} \tag{21}$$

Note: Because we set $s_0 = 0$ for identifiability, the term $f(0, h_0(\mathbf{z}))$ will appear with the same exponent k_0 in the numerator and denominator of (19). Therefore, the conditional likelihood does not identify $f(0,t)$. The separability condition (20) is required only for $y \geq 1$. Without loss of generality, we may impose the same factorization on $f(0,t)$.

Proof. To prove sufficiency, assume f factors as in (20). Then the conditional likelihood (19) simplifies as

$$\begin{aligned}
CL(\mathbf{h}, \mathbf{s}, \boldsymbol{\lambda} \mid \mathbf{k}) &= \frac{\prod_{i=1}^{m} g(y_i, h_{y_i}(\mathbf{z}))G(y_i, s_{y_i}\lambda_i)}{\sum_{\mathbf{y}^* \in \mathcal{K}} \prod_{i=1}^{m} g(y_i^*, h_{y_i^*}(\mathbf{z}))G(y_i^*, s_{y_i^*}\lambda_i)} && (22)\\
&= \frac{\prod_{i=1}^{m} g(y_i, h_{y_i}(\mathbf{z})) \prod_{i=1}^{m} G(y_i, s_{y_i}\lambda_i)}{\sum_{\mathbf{y}^* \in \mathcal{K}} \prod_{i=1}^{m} g(y_i^*, h_{y_i^*}(\mathbf{z})) \prod_{i=1}^{m} G(y_i^*, s_{y_i^*}\lambda_i)} && (23)\\
&= \frac{\prod_{j=0}^{J} g(j, h_j(\mathbf{z}))^{k_j} \prod_{i=1}^{m} G(y_i, s_{y_i}\lambda_i)}{\sum_{\mathbf{y}^* \in \mathcal{K}} \prod_{j=0}^{J} g(j, h_j(\mathbf{z}))^{k_j} \prod_{i=1}^{m} G(y_i^*, s_{y_i^*}\lambda_i)} && (24)\\
&= \frac{\prod_{i=1}^{m} G(y_i, s_{y_i}\lambda_i)}{\sum_{\mathbf{y}^* \in \mathcal{K}} \prod_{i=1}^{m} G(y_i^*, s_{y_i^*}\lambda_i)} && (25)
\end{aligned}$$

where g loses its dependence on y_i and y_i^* in (24) because the value of the products of $g(y_i, h_{y_i}(\mathbf{z}))$ and $g(y_i^*, h_{y_i^*}(\mathbf{z}))$ do not depend on the ordering of y_i and y_i^*. The y_i^* in the denominator are the same numerical values as the y_i in the numerator, only rearranged by the permutations enumerated in $\mathcal{K}$. As a result, $\mathbf{y}$ and every $\mathbf{y}^* \in \mathcal{K}$ share the same count vector $\mathbf{k}$. Therefore,

$$\prod_{i=1}^{m} g\big(y_i, h_{y_i}(\mathbf{z})\big) = \prod_{i=1}^{m} g\big(y_i^*, h_{y_i^*}(\mathbf{z})\big) = \prod_{j=0}^{J} g\big(j, h_j(\mathbf{z})\big)^{k_j} \tag{26}$$

allowing the substitution in (24). As a result, g drops out of the conditional likelihood and only G remains in (25) proving sufficiency.

To prove necessity, let $u = h_1(\mathbf{z})$ and $\mathbf{y} = [1, 0, 0, \ldots, 0]$. The choice of setting $y_1 = 1$ is arbitrary and, more generally, the subsequent arguments still follow if any one $y_i \neq 0$. With this setting, the conditional likelihood (19) as a function of u is

$$
\begin{aligned}
CL(u) &= \frac{f(1, u + s_1\lambda_1) \prod_{i=2}^{m} f(0, h_0(\mathbf{z}) + s_0\lambda_i)}{\sum_{I=1}^{m} \left(f(1, u + s_1\lambda_I) \prod_{i \neq I} f(0, h_0(\mathbf{z}) + s_0\lambda_i) \right)} && (27) \\
&= \frac{f(1, u + \lambda_1) \prod_{i=2}^{m} f(0, h_0(\mathbf{z}))}{\sum_{I=1}^{m} \left(f(1, u + \lambda_I) \prod_{i \neq I} f(0, h_0(\mathbf{z})) \right)} && (28) \\
&= \frac{(f(0, h_0(\mathbf{z})))^{m-1} f(1, u + \lambda_1)}{(f(0, h_0(\mathbf{z})))^{m-1} \sum_{I=1}^{m} f(1, u + \lambda_I)} && (29) \\
&= \frac{f(1, u + \lambda_1)}{\sum_{I=1}^{m} f(1, u + \lambda_I)} && (30)
\end{aligned}
$$

where (28) inserts $s_0 = 0$ and $s_1 = 1$ for identifiability. I in the denominator iterates through setting a single element of $\mathbf{y}^*$ to 1. Since $u = h_1(\mathbf{z})$, for the conditional likelihood to be independent of h, $CL(u)$ must be a constant, implying that $\frac{\partial}{\partial u} CL(u) = 0$.

$$
\begin{aligned}
0 &= \frac{\partial}{\partial u} CL(u) \\
&= \frac{\left(\sum_{I=1}^{m} f(1, u + \lambda_I) \right) f'(1, u + \lambda_1) - \left(\sum_{I=1}^{m} f'(1, u + \lambda_I) \right) f(1, u + \lambda_1)}{\left(\sum_{I=1}^{m} f(1, u + \lambda_I) \right)^2} && (31)
\end{aligned}
$$

Simplifying yields

$$
\frac{f'(1, u + \lambda_1)}{f(1, u + \lambda_1)} = \frac{\sum_{I=1}^{m} f'(1, u + \lambda_I)}{\sum_{I=1}^{m} f(1, u + \lambda_I)} \tag{32}
$$

Let $t = u + \lambda_1$. Then

$$
\frac{f'(1, t)}{f(1, t)} = \frac{\sum_{I=1}^{m} f'(1, t + \lambda_I - \lambda_1)}{\sum_{I=1}^{m} f(1, t + \lambda_I - \lambda_1)} \tag{33}
$$

If we substitute $t \leftarrow t + \lambda_1 - \lambda_L$ for some $2 \leq L \leq m$, then

$$
\frac{f'(1, t + \lambda_1 - \lambda_L)}{f(1, t + \lambda_1 - \lambda_L)} = \frac{\sum_{I=1}^{m} f'(1, t + \lambda_I - \lambda_L)}{\sum_{I=1}^{m} f(1, t + \lambda_I - \lambda_L)} \tag{34}
$$

We arrived at (33) assuming that $y_1 = 1$ and $y_i = 0$ for $i \neq 1$. If instead we had set $y_L = 1$ and $y_i = 0$ for $i \neq L$ then we would have arrived at

$$
\frac{f'(1, t)}{f(1, t)} = \frac{\sum_{I=1}^{m} f'(1, t + \lambda_I - \lambda_L)}{\sum_{I=1}^{m} f(1, t + \lambda_I - \lambda_L)} \tag{35}
$$

Together (34) and (35) imply that

$$
\frac{f'(1, t)}{f(1, t)} = \frac{f'(1, t + \lambda_1 - \lambda_L)}{f(1, t + \lambda_1 - \lambda_L)} \tag{36}
$$

showing that $\frac{f'(1,t)}{f(1,t)}$ is a periodic function with periodicity $\lambda_1 - \lambda_L$. Since (34) and (35) hold for every choice for $\boldsymbol{\lambda}$, choose λ_{L_1} and λ_{L_2} such that $\frac{\lambda_1 - \lambda_{L_1}}{\lambda_1 - \lambda_{L_2}} \notin \mathbb{Q}$, giving two incommensurate periods. Since $\frac{f'(1,t)}{f(1,t)}$ is a continuous function with two incommensurate periods (due to the arbitrary choice of $\boldsymbol{\lambda}$), it must be a constant. While derived here assuming $y_1 = 1$, the same derivation applies for $y_1 \in \{2, \ldots, J\}$. So for $y \geq 1$, $\frac{f'(y,t)}{f(y,t)}$ must be a constant.

$$
\begin{aligned}
\frac{f'(y, t)}{f(y, t)} &= c_{0y} && (37) \\
\frac{\partial}{\partial t} \log f(y, t) &= c_{0y} && (38) \\
\log f(y, t) &= c_{0y} t + c_{1y} && (39) \\
f(y, t) &= \exp\left(c_{0y} t + c_{1y} \right) && (40)
\end{aligned}
$$

Setting $t = h_y(\mathbf{z}) + s_y\lambda$ yields

$$
\begin{aligned}
f(y, h_y(\mathbf{z}) + s_y\lambda) &= \exp\left(c_{0y}(h_y(\mathbf{z}) + s_y\lambda) + c_{1y}\right) && (41)\\
&= e^{c_{1y}} \exp\left(c_{0y}h_y(\mathbf{z})\right) \exp\left(c_{0y}s_y\lambda\right) && (42)\\
&= g(y, h_y(\mathbf{z}))G(y, s_y\lambda) && (43)
\end{aligned}
$$

proving the factorization in (20) is also necessary. □

Theorem 4 showed that f must factor into two terms that separate h and $s\lambda$. It also signaled that both g and G must have an exponential form. Theorem 5 shows that the specific form is a multinomial model. Here $\lambda_y(\mathbf{x})$ is a regression function with officer-level covariates $\mathbf{x}$, a more general specification than in the main text of the article where $\lambda_y(\mathbf{x}_i) = s_y\lambda_i$, a scaled officer fixed effect.

Theorem 5: Uniqueness of multinomial logistic model for eliminating matched features in a conditional likelihood

If Y is a random variable for which the following three properties hold

1. $P(Y = y|\mathbf{z}, \mathbf{x}) = f(y, h_y(\mathbf{z}) + \lambda_y(\mathbf{x}))$ where $y \in \mathbb{N}_0$, $h_y(\mathbf{z}) + \lambda_y(\mathbf{x}) \in \mathbb{R}$, and
2. $f(y, t) : (\mathbb{N}_0, \mathbb{R}) \to [0, 1]$ is a continuous function of t, and
3. $f\big(y, h + \lambda\big) = g(y, h)\, G(y, \lambda)$, where $g, G : (\mathbb{N}_0, \mathbb{R}) \to \mathbb{R}_+$

then Y has a multinomial distribution with

$$
f\big(y, h_y(\mathbf{z}) + \lambda_y(\mathbf{x})\big) = \frac{e^{h_y^0(\mathbf{z})+\lambda_y^0(\mathbf{x})}}{\sum_{j=0}^{J} e^{h_j^0(\mathbf{z})+\lambda_j^0(\mathbf{x})}} \tag{44}
$$

where $h_y^0(\mathbf{z})$ and $\lambda_y^0(\mathbf{x})$ are linear transformations of $h_y(\mathbf{z})$ and $\lambda_y(\mathbf{x})$

Proof. For a fixed value of y,

$$
f(y, h_y(\mathbf{z}) + \lambda_y(\mathbf{x})) = g(y, h_y(\mathbf{z}))G(y, \lambda_y(\mathbf{x})) \tag{45}
$$

has the form of a multiplicative Pexider functional equation in $h_y(\mathbf{z})$ and $\lambda_y(\mathbf{x})$ (Aczél, 1966, §3.1, Thm. 2, page 143). The (non-trivial) continuous solution to this functional equation requires both g and G to be exponential functions so that

$$
f(y, h_y(\mathbf{z}) + \lambda_y(\mathbf{x})) = ce^{a_y + b_y(h_y(\mathbf{z}) + \lambda_y(\mathbf{x}))} \tag{46}
$$

where c, a_y, and b_y are real valued constants.

Since f represents a discrete probability function, introduce the normalizing denominator so that the sum of f over values of y is 1. The constant c cancels out in the normalization.

$$
f(y, h_y(\mathbf{z}) + \lambda_y(\mathbf{x})) = \frac{e^{a_y + b_y(h_y(\mathbf{z}) + \lambda_y(\mathbf{x}))}}{\sum_{j=0}^{J} e^{a_j + b_j(h_j(\mathbf{z}) + \lambda_j(\mathbf{x}))}} \tag{47}
$$

Since a_y and b_y are not identifiable separate from h_y and λ_y, without loss of generality, absorb them into $h_y^0(\mathbf{z}) = a_y + b_y h_y(\mathbf{z})$ and $\lambda_y^0(\mathbf{x}) = b_y\lambda_y(\mathbf{x})$

$$
f(y, h_y(\mathbf{z}) + \lambda_y(\mathbf{x})) = \frac{e^{h_y^0(\mathbf{z})+\lambda_y^0(\mathbf{x})}}{\sum_{j=0}^{J} e^{h_j^0(\mathbf{z})+\lambda_j^0(\mathbf{x})}} \tag{48}
$$

which has the familiar multinomial logistic regression model form. □

Theorem 5 implies that the ordinal model must have the form

$$f(y, h_y(\mathbf{z}) + s_y\lambda_i) = \frac{e^{h_y(\mathbf{z})+s_y\lambda_i}}{\sum_{j=0}^{J} e^{h_j(\mathbf{z})+s_j\lambda_i}} \tag{49}$$

The requirement that relative risk increases with λ imposes one additional constraint.

Theorem 6: Increasing relative risk implies ordering of s (Anderson's ordinal stereotype model)

If

$$\frac{P(Y = y_2|\mathbf{z}, \lambda + \epsilon)}{P(Y = y_1|\mathbf{z}, \lambda + \epsilon)} > \frac{P(Y = y_2|\mathbf{z}, \lambda)}{P(Y = y_1|\mathbf{z}, \lambda)}, \quad y_2 > y_1, \epsilon > 0 \tag{50}$$

and

$$f(y, h_y(\mathbf{z}) + s_y\lambda_i) = \frac{e^{h_y(\mathbf{z})+s_y\lambda_i}}{\sum_{j=0}^{J} e^{h_j(\mathbf{z})+s_j\lambda_i}} \tag{51}$$

then

$$s_0 < s_1 < \ldots < s_J \tag{52}$$

Proof. Substituting (50) into (51) implies

$$\frac{e^{h_{y_2}(\mathbf{z})+s_{y_2}(\lambda+\epsilon)}}{e^{h_{y_1}(\mathbf{z})+s_{y_1}(\lambda+\epsilon)}} > \frac{e^{h_{y_2}(\mathbf{z})+s_{y_2}\lambda}}{e^{h_{y_1}(\mathbf{z})+s_{y_1}\lambda}} \tag{53}$$

$$e^{(s_{y_2}-s_{y_1})\epsilon} > 1 \tag{54}$$

$$s_{y_2} > s_{y_1} \tag{55}$$

Since this applies to all values $y_2 > y_1$, then $s_0 < s_1 < \ldots < s_J$ □

Therefore, Anderson's ordinal stereotype model is the only ordinal model with a conditional likelihood that is invariant to environmental features and has increasing risk of more severe force with greater λ.

References

Aczél, J. (1966). *Lectures on Functional Equations and Their Applications*, Volume 19. New York: Academic Press.

Anderson, J. A. (1984). Regression and ordered categorical variables (with discussion). *Journal of the Royal Statistical Society, Series B 46*(1), 1–30.

Ridgeway, G. (n.d.). A conditional ordinal stereotype model to estimate police officers' propensity to escalate force. *Journal of the American Statistical Association*, 1–20.