WORLD STATISTICS CONGRESS 2025 THE HAGUE

International Statistical Institute

WELCOME.

HITOTSUBASHI UNIVERSITY

International Statistical Institute

WORLD STATISTICS CONGRESS 2025 THE HAGUE

CPS 14
Ordinal Data and Tree-Based Methods

Ryoya Yamasaki
Hitotsubashi University

Wednesday 8 October, 4:00PM - 5:00PM

Research Subject: Ordinal Regression

Title is “Mixture-Based Approximately Unimodal Likelihood Model for Ordinal Data”.

Subject is prediction (ordinal regression; OR) for ordinal data $(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \cdots \sim (\boldsymbol{X}, Y)$:

- Estimation of $\Pr(y|\boldsymbol{x})$
- Classification

- Explanatory variable $\boldsymbol{X} \in \mathbb{R}^d$ can be in any format
- Target variable $Y \in [K] \coloneqq \{1, \dots, K\}$ is categorical and has a “**natural ordinal relation**” $1 \prec \cdots \prec K$

Many OR applications:
age estimation from facial image; analysis of rating in EC site, credit rating & questionnaire

Facial image $\boldsymbol{X} = \boldsymbol{x}$

Age $Y = y$ 20 $\prec$ 40

Aaa, Aa, A, Baa, Ba, B, Caa, Ca, C

strongly agree, ag., neutral, disag., strongly disag.

Mathematical characterization of “natural ordinal relation” is key for success in OR.

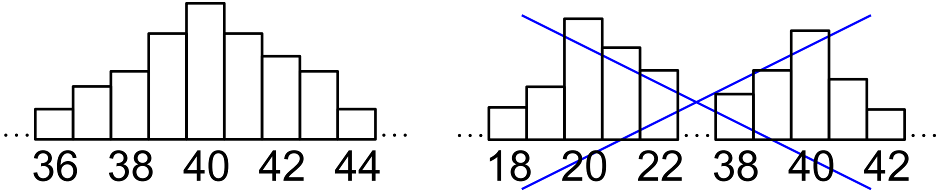
Previous Work (Yamasaki,22)

Previous works considered that **conditional probability distribution (CPD)** $(\Pr(y|\boldsymbol{x}))_{y\in[K]}$ **should be** **unimodal** **in a large domain of** X:

$$\Pr(1|\boldsymbol{x}) \le \cdots \le \Pr(M_{\boldsymbol{x}}|\boldsymbol{x}) \text{ and } \Pr(M_{\boldsymbol{x}}|\boldsymbol{x}) \ge \cdots \ge \Pr(K|\boldsymbol{x}) \text{ with a conditional mode } M_{\boldsymbol{x}} \in [K]$$

Facial image
$\boldsymbol{X} = \boldsymbol{x}$

CPD of age Y
$(\Pr(y|\boldsymbol{x}))_{y\in[K]}$



・J. F. P. da Costa+, "Classification of ordinal data using neural networks," *ECML*, 2005.
・M. Iannario+, "Cub models: statistical methods and empirical evidence," in *Modern Analysis of Customer Surveys*, 2011.
・C. Beckham+, "Unimodal probability distributions for deep ordinal classification," *ICML*, 2017.
・R. Yamasaki, "Unimodal likelihood models for ordinal data," *TMLR*, 2022.

High Unimodality Rate

Setting:

- **21 real-world ordinal datasets** used in a survey (Gutierrez+,15)
- **100 trials**, **training data size: 800** (large), validation/test data size: 100/remaining
- Estimate CPD $(\Pr(y|\boldsymbol{x}))_{y\in[K]}$ by **multinominal logit model** with a neural network
- Evaluate **unimodality rate (UR)** $\mathbb{E}\left[\mathbb{I}\{(\Pr(y|\boldsymbol{x}))_{y\in[K]} \text{ is unimodal.}\}\right]$ with test data

Result: **Many real-world ordinal data would have "high unimodality rate".**

Tab.1: **mean±std of UR** & evaluation for uniform random data (for comparison in same K)

dataset	d	K	UR
BA5'	32	5	.9999±.0006
SWD	10	4	.9996±.0031
WQR	11	6	.9959±.0186
CO5'	21	5	.9958±.0250
CO5	12	5	.9889±.0459
CE5'	16	5	.9814±.0366
BA5	8	5	.9760±.0894
LEV	4	5	.9594±.0570
CAR	21	4	.9380±.1693
CH5	8	5	.9334±.0924
CE5	8	5	.8866±.1102
BA10	8	10	.8831±.2141
AB5	10	5	.8785±.1024
CO10'	21	10	.8605±.1948
CO10	12	10	.8308±.1917
CE10'	16	10	.7239±.2580
ERA	4	9	.7122±.1581
BA10'	32	10	.6805±.3600
CH10	8	10	.5086±.3127
CE10	8	10	.4535±.2263
AB10	10	10	.3232±.1989

dataset	d	K	UR
uniform on Δ_3	-	4	.3326±.0135
〃 Δ_4	-	5	.1314±.0108
〃 Δ_5	-	6	.0443±.0065
〃 Δ_8	-	9	.0009±.0010
〃 Δ_9	-	10	.0001±.0003

Low: .0001~.3326

High: .3232~.9999

・P. A. Gutierrez et al., "Ordinal regression methods: survey and experimental study," *TKDE*, 2015.

Unimodal Likelihood (UL) Model

(Yamasaki,22) developed a UL model for modeling "high unimodality rate" data:

Multinominal logit model:

$P_{\mathrm{sm},y}(\boldsymbol{g}(\boldsymbol{x}))$

with

$\boldsymbol{g} : \mathbb{R}^d \to \mathbb{R}^K$

$$P_{\mathrm{sm},y}(\boldsymbol{u}) \equiv \frac{e^{u_y}}{\sum_{k=1}^{K} e^{u_k}}$$

UL model:

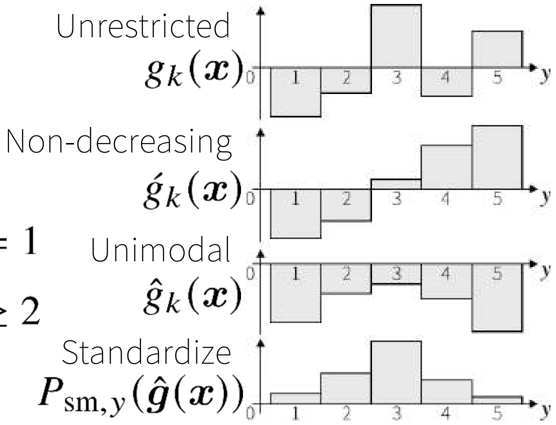
$P_{\mathrm{sm},y}(\hat{\boldsymbol{g}}(\boldsymbol{x}))$

with

$\boldsymbol{g} : \mathbb{R}^d \to \mathbb{R}^K$

$$\acute{g}_k(\boldsymbol{x}) \equiv \begin{cases} g_1(\boldsymbol{x}) & \text{for } k = 1 \\ \acute{g}_{k-1}(\boldsymbol{x}) + \{g_k(\boldsymbol{x})\}^2 & \text{for } k \geq 2 \end{cases}$$

$$\hat{g}_k(\boldsymbol{x}) \equiv -\{\acute{g}_k(\boldsymbol{x})\}^2 \quad \text{for } k = 1, \ldots, K$$



Theorem (simplified)

UL model is ensured to be unimodal & can represent any unimodal CPD:

$$\{(P_{\mathrm{sm},y}(\hat{\boldsymbol{g}}(\cdot)))_{y \in [K]} \mid \boldsymbol{g} : \mathbb{R}^d \to \mathbb{R}^K\} = \{\boldsymbol{p} : \mathbb{R}^d \to \hat{\Delta}_{K-1}\} \text{ with } \hat{\Delta}_{K-1} \equiv \{\text{unimodal } \boldsymbol{p} \in \Delta_{K-1}\}$$

・R. Yamasaki, "Unimodal likelihood models for ordinal data," *TMLR*, 2022.

This Work

Motivation:

In many ordinal data, CPD $(\Pr(y|x))_{y\in[K]}$ **would be non-unimodal at some points** $X = x$**:**

Tab.1: **mean±std of UR** & evaluation for uniform random data (for comparison in same K)

dataset	d	K	UR
BA5'	32	5	.9999±.0006
SWD	10	4	.9996±.0031
WQR	11	6	.9959±.0186
CO5'	21	5	.9958±.0250
CO5	12	5	.9889±.0459
CE5'	16	5	.9814±.0366
BA5	8	5	.9760±.0894
LEV	4	5	.9594±.0570
CAR	21	4	.9380±.1693
CH5	8	5	.9334±.0924
CE5	8	5	.8866±.1102
BA10	8	10	.8831±.2141
AB5	10	5	.8785±.1024
CO10'	21	10	.8605±.1948
CO10	12	10	.8308±.1917
CE10'	16	10	.7239±.2580
ERA	4	9	.7122±.1581
BA10'	32	10	.6805±.3600
CH10	8	10	.5086±.3127
CE10	8	10	.4535±.2263
AB10	10	10	.3232±.1989

dataset	d	K	UR
uniform on Δ_3	-	4	.3326±.0135
〃 Δ_4	-	5	.1314±.0108
〃 Δ_5	-	6	.0443±.0065
〃 Δ_8	-	9	.0009±.0010
〃 Δ_9	-	10	.0001±.0003

High but not 1

UL model has a bias at points with a non-unimodal CPD.

Goal:

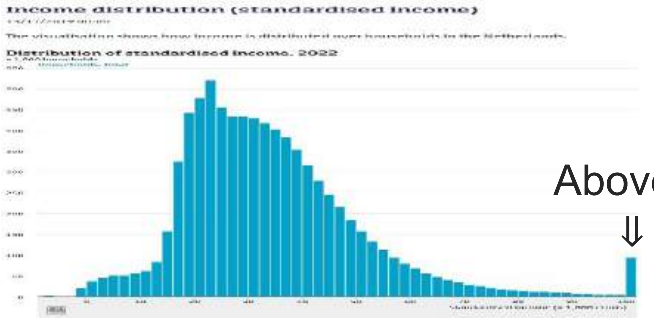
To decrease bias and improve prediction performance, develop a novel likelihood model with better representation of ordinal data (higher representation ability).

Working Hypothesis

We consider that CPD $(\Pr(y|\boldsymbol{x}))_{y\in[K]}$ **of ordinal data should be close to be unimodal at points** $X=\boldsymbol{x}$ **where** $(\Pr(y|\boldsymbol{x}))_{y\in[K]}$ **is not strictly unimodal.**

Rating distribution in EC site (amazon.com)

Frequency distribution of Household income (cbs.nl/en-gb/visualisations/income-distribution)



Above ~
⇓

Mathematically, using Hausdorff distance $D_H(\boldsymbol{p},S)\equiv\min_{\boldsymbol{q}\in S}||\boldsymbol{p}-\boldsymbol{q}||$ and unimodal set $\widehat{\Delta}_{K-1}\equiv\{\text{unimodal } \boldsymbol{p}\in\Delta_{K-1}\}$, **we further assume that unimodality deviation** $D_H\left((\Pr(y|\boldsymbol{x}))_{y\in[K]},\widehat{\Delta}_{K-1}\right)$ **of CPD** $(\Pr(y|\boldsymbol{x}))_{y\in[K]}$ **is low at most** $X=\boldsymbol{x}$.

Low Unimodality Deviation

Setting:
- **21 real-world ordinal datasets** used in a survey (Gutierrez+,15)
- **100 trials**, **training data size: 800** (large), validation/test data size: 100/remaining
- Estimate CPD $(\Pr(y|\boldsymbol{x}))_{y\in[K]}$ by **multinominal logit model** with a neural network
- Evaluate **mean unimodality deviation (MUD)** $\mathbb{E}\left[D_H\left((\Pr(y|\boldsymbol{x}))_{y\in[K]}, \widehat{\Delta}_{K-1}\right)\right]$ with test data

Result: **Many real-world ordinal data would have "low unimodality deviation".**

Tab.2: **mean±std of MUD** & evaluation for uniform random data (for comparison in same K)

dataset	d	K	MUD	dataset	d	K	MUD	dataset	d	K	MUD
BA5'	32	5	.0000±.0000	LEV	4	5	.0003±.0007	CO10	12	10	.0009±.0019
SWD	10	4	.0000±.0002	CAR	21	4	.0003±.0010	CE10'	16	10	.0017±.0023
WQR	11	6	.0000±.0001	CH5	8	5	.0009±.0017	ERA	4	9	.0048±.0049
CO5'	21	5	.0000±.0000	CE5	8	5	.0020±.0024	BA10'	32	10	.0012±.0018
CO5	12	5	.0001±.0003	BA10	8	10	.0000±.0001	CH10	8	10	.0061±.0072
CE5'	16	5	.0004±.0007	AB5	10	5	.0017±.0022	CE10	8	10	.0058±.0052
BA5	8	5	.0000±.0001	CO10'	21	10	.0003±.0009	AB10	10	10	.0105±.0086

dataset	d	K	MUD
uniform on Δ_3	-	4	.0752±.0026
〃 Δ_4	-	5	.1000±.0023
〃 Δ_5	-	6	.1162±.0023
〃 Δ_8	-	9	.1365±.0016
〃 Δ_9	-	10	.1385±.0014

High: .0752~.1385

Low: .0000~.0105

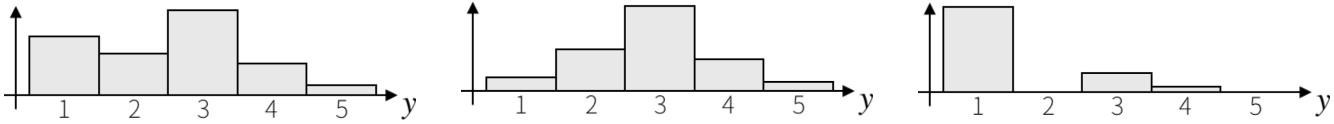
・P. A. Gutierrez et al., "Ordinal regression methods: survey and experimental study," *TKDE*, 2015.

Approximately Unimodal Likelihood (AUL) Model

We propose a mixture-based AUL model for modeling "low unimodality deviation" data :

$$P_{\mathrm{AUL},y}(\boldsymbol{g}_1(\boldsymbol{x}), \boldsymbol{g}_2(\boldsymbol{x}); r) \equiv (1-r)P_{\mathrm{sm},y}(\hat{\boldsymbol{g}}_1(\boldsymbol{x})) + rP_{\mathrm{sm},y}(\boldsymbol{g}_2(\boldsymbol{x}))$$ with $\boldsymbol{g}_1, \boldsymbol{g}_2 : \mathbb{R}^d \to \mathbb{R}^K$,

unimodal likelihood — multinominal logit — $r \in [0,1]$: **mixture rate**



Theorem (simplified)

1. **AUL model can represent any unimodal CPD:**
   $\{P_{\mathrm{AUL},y}(\boldsymbol{g}_1(\cdot), \boldsymbol{g}_2(\cdot); r) \mid \boldsymbol{g}_1, \boldsymbol{g}_2 : \mathbb{R}^d \to \mathbb{R}^K\} \supseteq \{\boldsymbol{p} : \mathbb{R}^d \to \hat{\Delta}_{K-1}\}$

   It can represent only unimodal or close to be unimodal CPD.

2. **Unimodality deviation of AUL model is bounded from above by $\sqrt{2}\,r$:**
   $D_{\mathrm{H}}((P_{\mathrm{AUL},y}(\boldsymbol{g}_1(\boldsymbol{x}), \boldsymbol{g}_2(\boldsymbol{x}); r))_{y\in[K]}, \hat{\Delta}_{K-1}) \le \sqrt{2}r$ for any $\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{x}$

3. **Representation ability of AUL model is non-decreasing in r:**
   $\{P_{\mathrm{AUL},y}(\boldsymbol{g}_1(\cdot), \boldsymbol{g}_2(\cdot); r_1) \mid \boldsymbol{g}_1, \boldsymbol{g}_2 : \mathbb{R}^d \to \mathbb{R}^K\} \subseteq \{P_{\mathrm{AUL},y}(\boldsymbol{g}_1(\cdot), \boldsymbol{g}_2(\cdot); r_2) \mid \boldsymbol{g}_1, \boldsymbol{g}_2 : \mathbb{R}^d \to \mathbb{R}^K\}$ if $r_1 \le r_2$

Experiment

Setting:

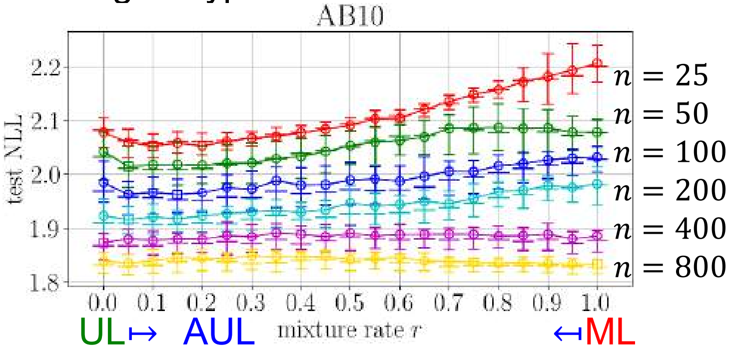
- **21 real-world ordinal datasets** used in a survey (Gutierrez+,15)
- **100 trials**, **training data size: $25, 50, \dots, 800$**, validation/test data size: 100/remaining
- Estimate CPD $(\Pr(y|\boldsymbol{x}))_{y\in[K]}$ by **multinominal logit (ML)** $(r=1)$, **unimodal likelihood (UL)** $(r=0)$, **approximately UL (AUL)** $(r \in \{0.05, 0.1, \dots, 0.95\})$ models with a neural network
- Evaluate **negative log likelihood (NLL)**; **mean zero-one & absolute error (MZE & MAE)** $\mathbb{E}[\ell(f(\boldsymbol{x}), y)]$, $\ell(j,k) = \mathbb{I}(j \neq k), |j-k|$ for a classifier $f(\boldsymbol{x}) = \mathrm{argmin}_{k\in[K]} \sum_{y=1}^{K} \ell(k,y) P_y(\boldsymbol{g}(\boldsymbol{x}))$

Result (NLL):

Tab.3: **Number of wins** over 21 datasets in Bonferroni correction with significance level 0.05 of U-test.

NLL	$n=25$	$n=50$	$n=100$	$n=200$	$n=400$	$n=800$
AUL vs ML	18,0	19,0	14,0	7,3	0,10	1,14
AUL vs UL	14,0	12,0	8,0	5,0	3,1	7,1

Fig.1: Typical behaviors in r & n



- P. A. Gutierrez et al., "Ordinal regression methods: survey and experimental study," *TKDE*, 2015.

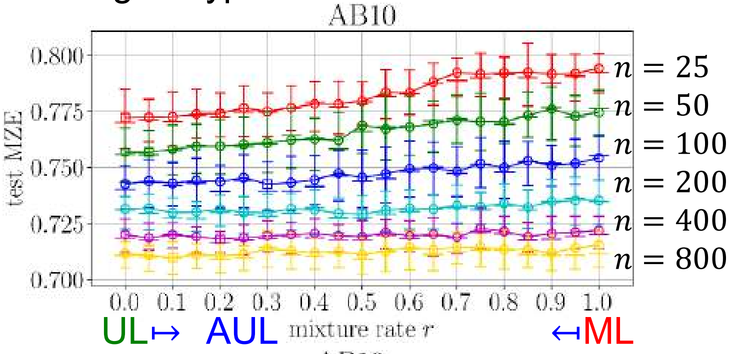
Result (MZE & MAE):

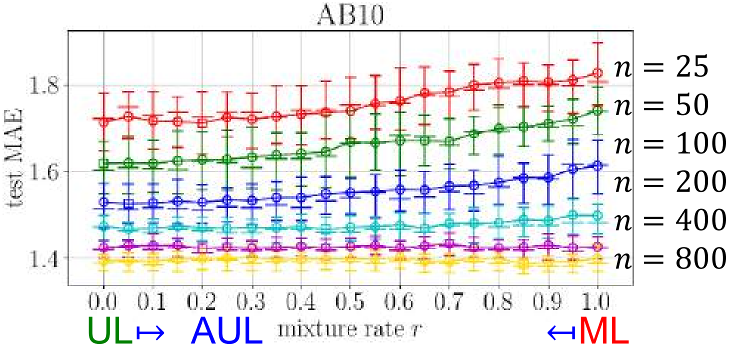
Tab.4: **Number of wins** over 21 datasets in Bonferroni correction with significance level 0.05 of U-test.

MZE	$n=25$	$n=50$	$n=100$	$n=200$	$n=400$	$n=800$
AUL vs ML	19,0	17,0	10,0	1,3	1,3	4,5
AUL vs UL	6,0	7,0	10,0	11,0	8,0	6,0

MAE	$n=25$	$n=50$	$n=100$	$n=200$	$n=400$	$n=800$
AUL vs ML	20,0	16,0	10,0	3,2	0,5	1,5
AUL vs UL	3,0	4,0	2,0	5,0	4,0	7,0

Fig.2: Typical behaviors in r & n





Regarding NLL, MZE & MAE,
AUL model was better than UL model,
and better than ML model for small n.

Conclusion

We verified that many ordinal data would have "low unimodality deviation".

Tab.2: **mean±std of MUD** & evaluation for uniform random data (for comparison in same K)

dataset	d	K	MUD
BA5'	32	5	.0000±.0000
SWD	10	4	.0000±.0002
WQR	11	6	.0000±.0001
CO5'	21	5	.0000±.0000
CO5	12	5	.0001±.0003
CE5'	16	5	.0004±.0007
BA5	8	5	.0000±.0001

dataset	d	K	MUD
LEV	4	5	.0003±.0007
CAR	21	4	.0003±.0010
CH5	8	5	.0009±.0017
CE5	8	5	.0020±.0024
BA10	8	10	.0000±.0001
AB5	10	5	.0017±.0022
CO10'	21	10	.0003±.0009

dataset	d	K	MUD
CO10	12	10	.0009±.0019
CE10'	16	10	.0017±.0023
ERA	4	9	.0048±.0049
BA10'	32	10	.0012±.0018
CH10	8	10	.0061±.0072
CE10	8	10	.0058±.0052
AB10	10	10	.0105±.0086

dataset	d	K	MUD
uniform on Δ_3	-	4	.0752±.0026
〃 Δ_4	-	5	.1000±.0023
〃 Δ_5	-	6	.1162±.0023
〃 Δ_8	-	9	.1365±.0016
〃 Δ_9	-	10	.1385±.0014

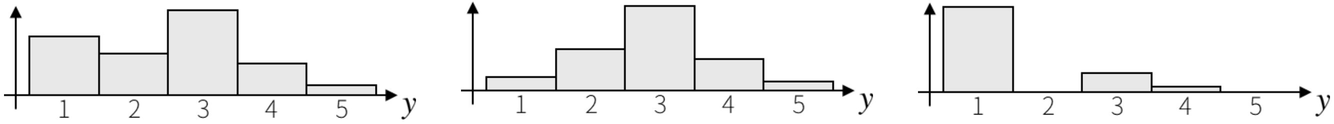
High: .0752~.1385

Low: .0000~.0105

We proposed a mixture-based approximately unimodal likelihood (AUL) model:

$$P_{\mathrm{AUL},y}(\boldsymbol{g}_1(\boldsymbol{x}),\boldsymbol{g}_2(\boldsymbol{x});r) \equiv (1-r)P_{\mathrm{sm},y}(\hat{\boldsymbol{g}}_1(\boldsymbol{x})) + rP_{\mathrm{sm},y}(\boldsymbol{g}_2(\boldsymbol{x})) \text{ with } \boldsymbol{g}_1,\boldsymbol{g}_2:\mathbb{R}^d\to\mathbb{R}^K,$$

unimodal likelihood multinominal logit $r \in [0,1]$: mixture rate



AUL model gave good prediction for ordinal data when training data size was small.

Thank you for listening! Contact: ryoya.yamasaki@r.hit-u.ac.jp

WORLD STATISTICS CONGRESS 2025 THE HAGUE

International Statistical Institute

APPENDIX

Representation ability

We define "representation ability" of a likelihood model as a functional of the CPD that that likelihood model can represent (see also (Yamasaki,22)):

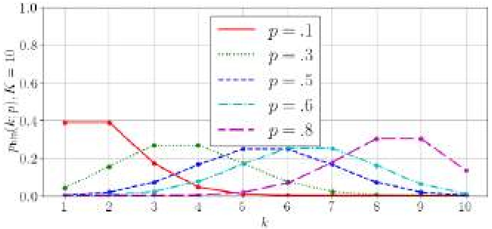
For likelihood model $\left(P_y(\boldsymbol{g}(\boldsymbol{x}))\right)_{y\in[K]}, \boldsymbol{g} \in G$, representation ability is $\left\{\left(P_y(\boldsymbol{g}(\boldsymbol{x}))\right)_{y\in[K]} \,\middle|\, \boldsymbol{g} \in G\right\}$.

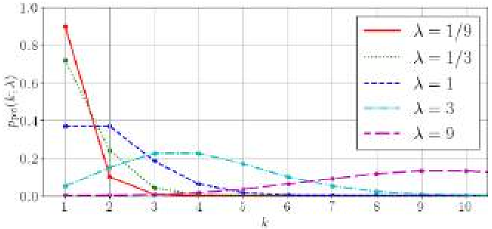
・R. Yamasaki, "Unimodal likelihood models for ordinal data," *TMLR*, 2022.

Other Unimodal Likelihood Models

(da Costa+,05) inspired by shape of binomial and Poisson distributions and developed

$$p_{\mathrm{bin}}(y, \tfrac{1}{1+e^{-g(\boldsymbol{x})}}) \text{ with } p_{\mathrm{bin}}(k,p) = \binom{K-1}{k-1} p^{k-1}(1-p)^{K-k} \qquad \frac{p_{\mathrm{poi}}(y, e^{g(\boldsymbol{x})})}{\sum_{k=1}^{K} p_{\mathrm{poi}}(k, e^{g(\boldsymbol{x})})} \text{ with } p_{\mathrm{poi}}(k,\lambda) = \frac{\lambda^{k-1}e^{-\lambda}}{(k-1)!}$$



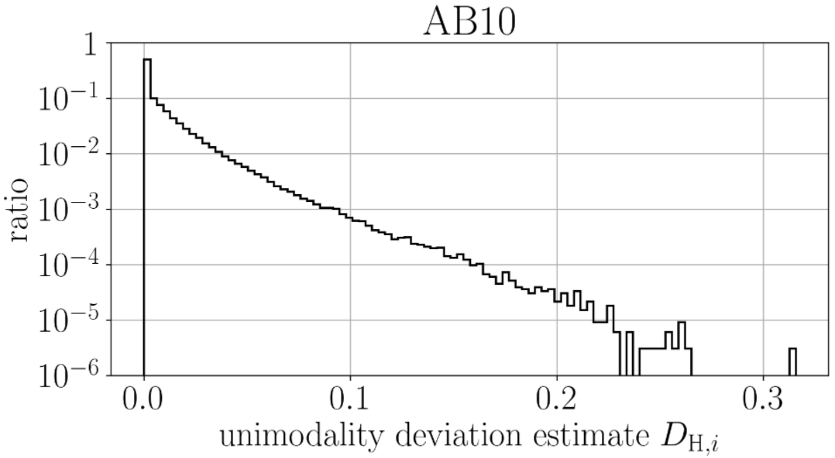


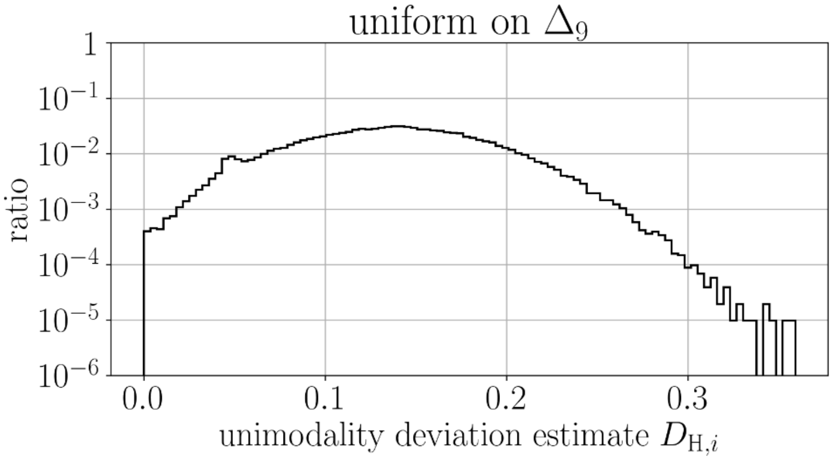
(Iannario+,11) developed a mixture of (da Costa+,05)'s binomial-based and uniform models,
(Beckham+,17) developed scaling-extension of (da Costa+,05),
(Yamasaki,22) developed various models with higher representation ability.

・J. F. P. da Costa+, "Classification of ordinal data using neural networks," *ECML*, 2005.
・M. Iannario+, "Cub models: statistical methods and empirical evidence," in *Modern Analysis of Customer Surveys*, 2011.
・C. Beckham+, "Unimodal probability distributions for deep ordinal classification," *ICML*, 2017.
・R. Yamasaki, "Unimodal likelihood models for ordinal data," *TMLR*, 2022.

Low Unimodality Deviation

We also evaluated distribution of unimodality deviation $D_H\left((\Pr(y|\boldsymbol{x}))_{y\in[K]}, \widehat{\Delta}_{K-1}\right)$ as well.

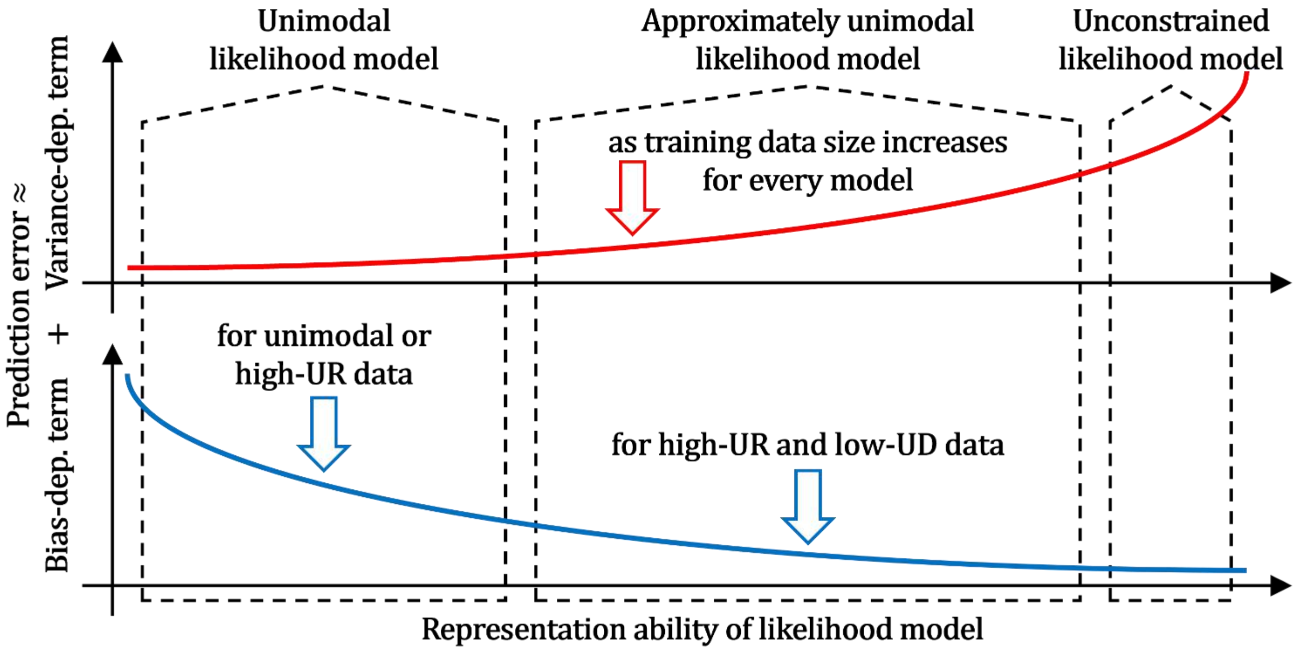




Regarding distribution, many real-world ordinal data would have “low unimodality deviation”.

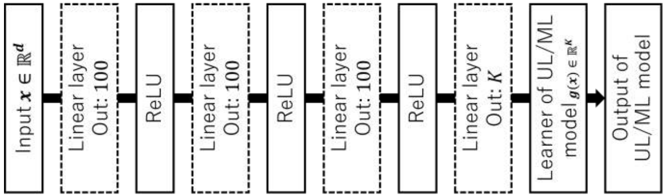
Expectation

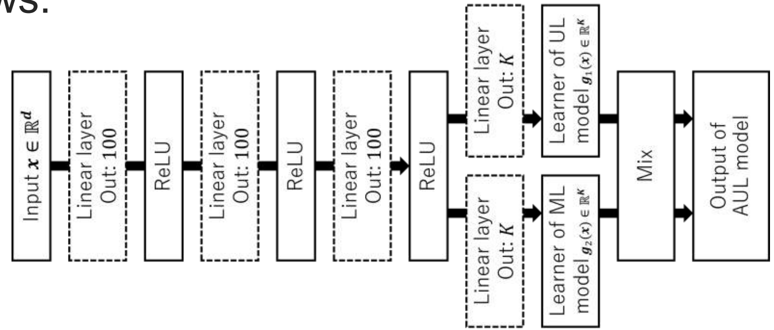
From bias-variance tradeoff, we expect that likelihood models with good representation of data and smaller representation ability will give good prediction performance with small-size training data.



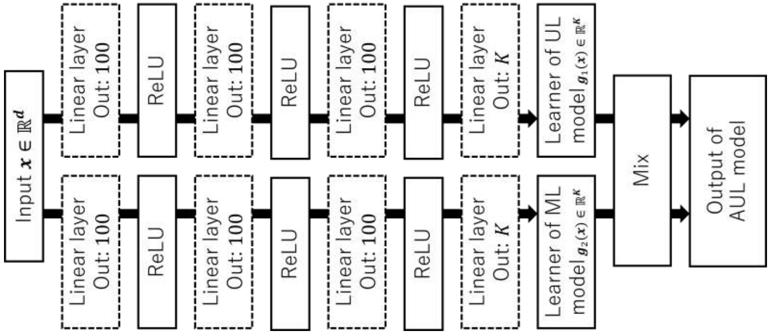
Setting of Experiment

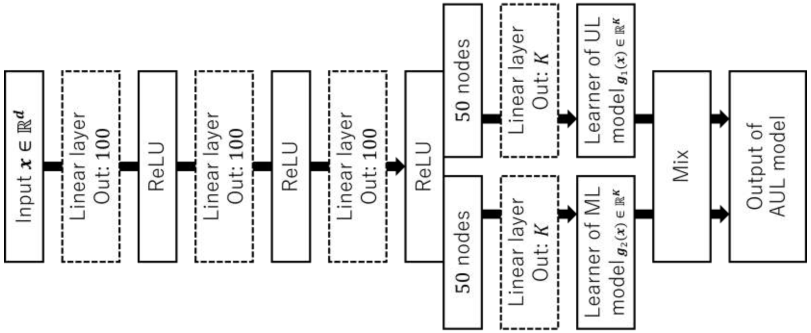
UL/ML and AUL models are implemented as follows:





Those models have different numbers of parameters.
We also tried different implementation of AUL model:





Setting of Experiment

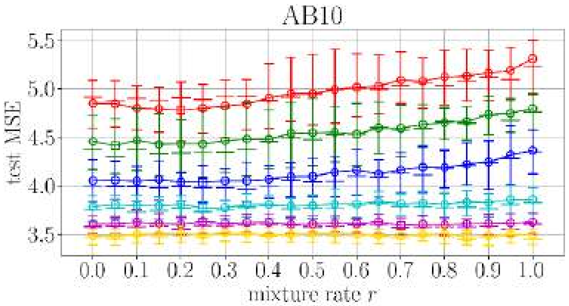
We trained every model by Adam optimization with the NLL as the objective function, mini-batch size 16, and ascending learning rate $10^{2t/300-4}$ at t-th epoch for 300 epochs.

At the end of each epoch, we evaluated the errors, NLL, MZE, MAE, and MSE, with validation data of the size 100.

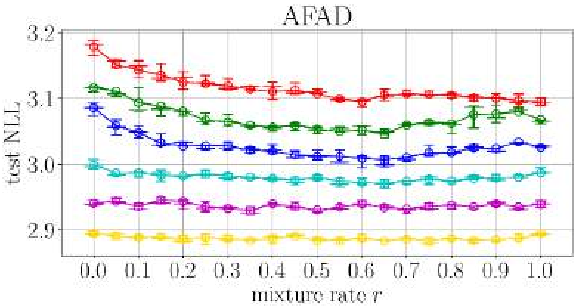
We then adopted a model at the epoch when the validation error got minimum (and simultaneously selected the mixture rate r), and calculated the error with remaining test data of the size $(n_{\mathrm{tot}} - n - 100)$ for that model.

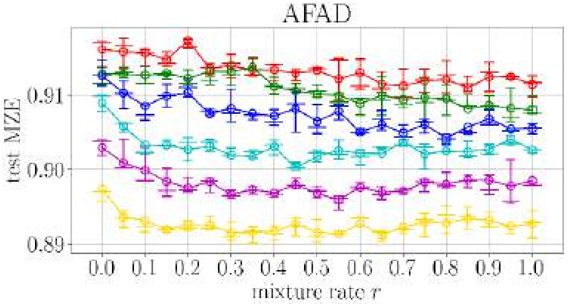
Result of Experiment

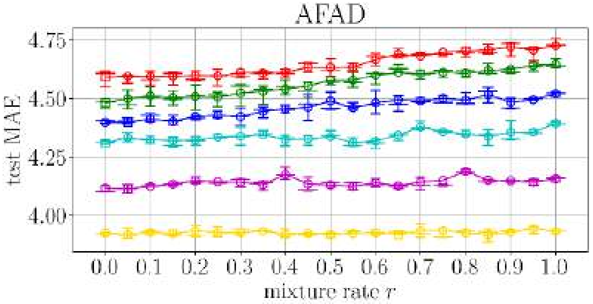
We also evaluate mean squares error (MSE).

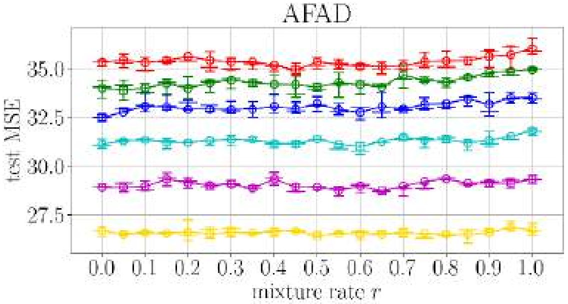


We also experimented face-age estimation with $n = 1250{,}2500{,}5000{,}10000{,}20000{,}40000$.









Comparison with Other Methods

We also experimented cumulative-logits (CL), proportional-odds (PO) CL, POUL models:

CL:

$$P_{\mathrm{cl},y}(\boldsymbol{g}(\boldsymbol{x})) \equiv \begin{cases} \frac{1}{1+e^{-g_1(\boldsymbol{x})}} & \text{for } k=1 \\ 1-\frac{1}{1+e^{-g_{K-1}(\boldsymbol{x})}} & \text{for } k=K \\ \frac{1}{1+e^{-g_y(\boldsymbol{x})}}-\frac{1}{1+e^{-g_{y-1}(\boldsymbol{x})}} & \text{for other } k \end{cases}$$

POCL:

$$P_{\mathrm{cl},y}(\boldsymbol{g}(\boldsymbol{x})) \text{ with } \boldsymbol{g}(\boldsymbol{x}) = (\acute{b}_1 - a(\boldsymbol{x}), \ldots, \acute{b}_{K-1} - a(\boldsymbol{x})),$$

$$\acute{b}_k = \begin{cases} b_1 & \text{for } k=1 \\ \acute{b}_{k-1} + \{b_k\}^2 & \text{for } k \geq 2 \end{cases}$$

POUL:

$$P_{\mathrm{sm},y}(\hat{\boldsymbol{g}}(\boldsymbol{x})) \text{ with } \hat{\boldsymbol{g}}(\boldsymbol{x}) = (-\{\acute{b}_1 - a(\boldsymbol{x})\}^2, \ldots, -\{\acute{b}_K - a(\boldsymbol{x})\}^2)$$

NLL	$n=25$	$n=50$	$n=100$	$n=200$	$n=400$	$n=800$
AUL vs ML	18,0	19,0	14,0	7,3	0,10	1,14
AUL vs UL	14,0	12,0	8,0	5,0	3,1	7,1
AUL vs CL	17,0	14,0	10,2	6,6	3,5	3,8
AUL vs POCL	16,1	19,0	16,1	14,4	12,7	7,4
AUL vs POUL	11,1	14,3	12,2	12,1	11,0	11,0

POCL model is popular and often called “ordinal regression model”.

AUL model is better for small n.

MZE	$n=25$	$n=50$	$n=100$	$n=200$	$n=400$	$n=800$
AUL vs ML	19,0	17,0	10,0	1,3	1,3	4,5
AUL vs UL	6,0	7,0	10,0	11,0	8,0	6,0
AUL vs CL	16,0	8,0	5,0	3,1	2,1	2,3
AUL vs POCL	18,1	17,0	16,0	15,0	14,1	10,2
AUL vs POUL	4,1	6,0	10,0	14,0	17,0	14,0

MAE	$n=25$	$n=50$	$n=100$	$n=200$	$n=400$	$n=800$
AUL vs ML	20,0	16,0	10,0	3,2	0,5	1,5
AUL vs UL	3,0	4,0	2,0	5,0	4,0	7,0
AUL vs CL	11,0	11,0	7,0	2,0	1,1	1,1
AUL vs POCL	6,0	8,0	12,0	8,0	9,1	6,2
AUL vs POUL	6,0	4,0	5,1	8,1	9,0	11,0

Comparison with Other Methods

We also experimented unimodality-promoting regularization learning (UPRL) methods.
Here, ML’ is ML + UPRL; AUL’ is AUL + UPRL.

NLL	$n=25$	$n=50$	$n=100$	$n=200$	$n=400$	$n=800$
AUL vs ML	18,0	19,0	14,0	7,3	0,10	1,14
AUL vs UL	14,0	12,0	8,0	5,0	3,1	7,1
AUL vs ML’	13,2	12,0	7,2	2,5	0,12	1,10
AUL vs AUL’	0,6	0,2	0,2	0,0	0,0	0,0

AUL model is better than ML’ for small n.
AUL’ model is the best for small n.

MZE	$n=25$	$n=50$	$n=100$	$n=200$	$n=400$	$n=800$
AUL vs ML	19,0	17,0	10,0	1,3	1,3	4,5
AUL vs UL	6,0	7,0	10,0	11,0	8,0	6,0
AUL vs ML’	15,0	7,0	3,0	0,1	2,3	4,4
AUL vs AUL’	0,2	0,0	1,0	1,0	0,0	1,1

MAE	$n=25$	$n=50$	$n=100$	$n=200$	$n=400$	$n=800$
AUL vs ML	20,0	16,0	10,0	3,2	0,5	1,5
AUL vs UL	3,0	4,0	2,0	5,0	4,0	7,0
AUL vs ML’	14,0	8,0	3,0	0,1	0,3	4,5
AUL vs AUL’	0,0	0,0	0,0	0,0	0,0	0,0

▪S. Belharbi+, “Non-parametric uni-modality constraints for deep ordinal classification,” *arXiv preprint arXiv:1911.10720*, 2020.
▪T. Albuquerque+, “Ordinal losses for classification of cervical cancer risk,” *PeerJ Computer Science*, vol. 7, p. e457, 2021.

THANK YOU.