Reliability for Binary and Ordinal Data in Forensics

Hina Arora, Naomi Kaplan-Damary, Hal Stern

Background

- Decisions in forensic examination procedure are often based on subjective opinion of an expert.
- Subjective decisions are prone to variation. Reliability and accuracy for such comparisons is usually evaluated using black box studies.
- Increasing interest in data that is on an ordinal scale.

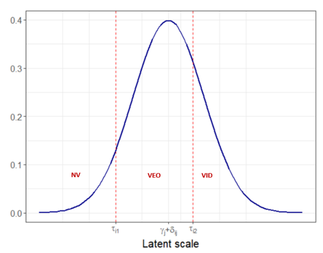
Objectives

- Provide a statistical model for assessing reliability of feature-based comparisons in forensic science.
- We also consider a mixture model approach that clusters examiners together based on their tendencies to rate samples.

Modelling Ordinal Data

- Consider a case where decisions can be on an ordinal scale of $1-3$ (e.g., no value, VEO, VID).
- Let Y_{ijk} be the ordinal decision by examiner i on sample j at the k^{th} trial, k allows for repeated observations.
- We assume that ordinal decisions are dependent on an underlying continuous variable Z_{ijk}.
- Assume $\tau_{i,1}$ and $\tau_{i,2}$ are examiner "thresholds" for classifying sample i into one of the three categories.

$$Y_{ijk} = 1 + I(Z_{ij} > \tau_{i,1}) + I(Z_{ij} > \tau_{i,2})$$
$$Z_{ijk}|\gamma_j, \delta_{ij} \sim N(\gamma_j + \delta_{ij}, 1)$$
$$\tau_{i,2} = \tau_{i,1} + \tau_0$$



Results with Fingerprints

- Thresholds of 169 examiners for the Analysis phase of the fingerprint analysis procedure.
- The data is on a scale of NV, VEO, and VID which indicate increasing level of detail in the latent print that makes it suitable for comparisons.
- The reproducibility was estimated to be 0.72 with a 95% credible interval of (0.69, 0.74) and the repeatability was 0.77 with a 95% credible interval (0.75, 0.79).

	% NV decisions	Estimated thresholds	Average estimated γ_j
Examiner A	10.8	-1.79	0.23
Examiner B	18.7	-1.79	0.09

DPM on Examiner Tendencies

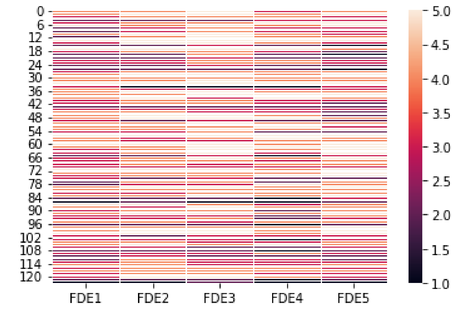
- We are interested in models that encourage parameter sharing. Consider DPMM.
- We will assume that there are unknown number of clusters of examiners such that examiners within a cluster share tendencies to rate samples.
- Assume that $\alpha_{c(i)}$ in the model follows a Dirichlet process mixture. Say that the base distribution is a normal distribution with a zero mean and certain variance σ_b^2.
- In this case, we assume that the cutpoints or the thresholds are shared between all examiners.

$$Y_{ijk} = 1 + I(Z_{ijk} > \kappa_1) + I(Z_{ijk} > \kappa_2)$$
$$\kappa_2 > \kappa_1$$
$$Z_{ijk} \sim N(\alpha_{c(i)} + \gamma_j, 1)$$

- The concentration parameter is sampled using adaptive rejection sampling.
- Similar to the above setup, sample difficulties could also be clustered.

Results with Signatures

- We analyzed the data collected from a handwritten signature complexity study.
- Five forensic document examiners provided complexity assessments for 123 signatures on a scale of 1-5.



- We found that there are two clusters of examiners based on examiner tendencies to rate samples.
- We found that examiners 1,4 were in the same cluster based on their tendencies. Examiners 2,3 were in another cluster.
- Note that we plan to look at studies with binary and ordinal data and more examiners.

Conclusions

- We provided a methodology for assessing examiner thresholds with ordinal decisions.
- We also provided a method to cluster examiners in groups in the absence of examiner covariates while accounting for sample difficulties.
- We were able to obtain insights into the data collected from two black-box studies.

References

- Stern, Hal S., et al. Assessing the complexity of handwritten signatures. *Law, Probability and Risk* 17.2 (2018): 123-132.
- Ulery, Bradford T., et al. Accuracy and reliability of forensic latent fingerprint decisions. *Proceedings of the National Academy of Sciences* 108.19 (2011): 7733-7738.
- Escobar, M. D., & West, M. (1995). "Bayesian density estimation and inference using mixtures." Journal of the american statistical association, 90(430), 577-588.
- Johnson, Valen E., and James H. Albert. *Ordinal data modeling*. Springer Science & Business Media, 2006.
- Stern, Hal S., Maria Cuellar, and David Kaye. Reliability and validity of forensic science evidence. *Significance* 16.2 (2019): 21-24.
- Albert, James H., and Siddhartha Chib. "Bayesian analysis of binary and polychotomous response data." *Journal of the American statistical Association* 88.422 (1993): 669-679.

This work was partially funded by CSAFE through Cooperative Agreement # 70NANB15H176 between NIST and Iowa State University, which includes activities carried out at Carnegie Mellon University, University of California Irvine, University of Virginia and Duke University.