OTTAWA 2023

64TH WORLD STATISTICS CONGRESS

Munich Center for Machine Learning

LMU

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

CPS 82 – Statistical Methodology I

A general framework for reporting methods in regression analysis

Hannah Kümpel, Sabine Hoffmann
LMU Munich

20 July 8:30 a.m. – 9:40 a.m.

Interpretable ML

"Determining trust in individual predictions is an important problem when [models are] used for decision making."
– Ribeiro (2016)

Medicine

"The reign of the p-value is over: what alternative analyses could we employ to fill the power vacuum?"
– Halsey (2019)

"Science is undergoing a crisis that has been referred to, since the early 21st century, as a crisis of confidence and a crisis of replication."
– Frias-Navarro (2020)

Sociology

"In disciplines with high theoretical diversity, such as sociology, [standard literature searches often do not suffice] to fully exploit the entire body of relevant work."
– Roelfs (2013)

Economicetrics

"Although the literature on how to implement and interpret interactions is vast, there remains confusion and disconsensus among researchers about how to interpret a main effect in the presence of moderation."
– Busenbark (2022)

Interpretable ML

- *Individual conditional expectation plots*
- *Marginal plots*
- *Partial dependence plots*

Medicine

- *Predicted change in probability*
- *Average treatment effect*

proposals for interpretable and comparable effect size measures and visualization techniques

Sociology

- *Average predictive comparisons*
- *Unit change*

Economicetrics

- *Marginal effects*
- *Adjusted predictions*

The framework's setting I

- The proposed framework applies to any (thus far parametric, i.e. finite dimensional parameter space Θ) models where the conditional expectation of the target variable can be written as

$$\mathbb{E}[Y|X] = g_\theta(X)$$

where the function $g_\theta : \mathbb{R}^p \longrightarrow \mathbb{R}$, $p \in \mathbb{N}$, is at least once partially differentiable w.r.t. each metric element of X $\forall \theta \in \Theta$.

- Here, we consider the function g_θ to be indexed by $\theta \in \Theta$ in the sense of the following mapping

$$g : \Theta \longrightarrow \mathcal{M}(\mathbb{R}^p), \quad \theta \longmapsto g_\theta$$

with $\mathcal{M}(\mathbb{R}^p) := \left\{f : \mathbb{R}^p \longrightarrow \mathbb{R} | \ f \text{ is } \left(\mathcal{B}(\mathbb{R}^p), \mathcal{B}(\mathbb{R})\right)\text{-measurable}\right\}$, $p \in \mathbb{N}$.

The framework's setting II

- Given this, we propose to utilize probability measures to derive appropriately weighted means of functions of g_θ over areas of interest.

- Specifically, for some function $h : \mathbb{R}^d \longrightarrow \mathbb{R}$, $d \in \mathbb{N}_{>0}$, and set $D \subseteq \mathbb{R}^{\tilde{d}}$, $d \geq \tilde{d} \in \mathbb{N}_{>0}$, we consider probability measures μ that satisfy the following requirements:

(M1) μ is a probability measure on $\left(\mathbb{R}^{\tilde{d}}, \mathcal{B}(\mathbb{R}^{\tilde{d}})\right)$.

(M2) $\text{supp}(\mu) \subseteq D$, if required as a result of μ being normalized w.r.t. D.

(M3) $\int_{\mathbb{R}^{\tilde{d}}} |h(x)| d\mu(x) < \infty$, if $\tilde{d} = d$, or $\forall x_a \in \mathbb{R}^{d-\tilde{d}} : \int_{\mathbb{R}^{\tilde{d}}} |h(x_a, x_b)| d\mu(x_b) < \infty$, if $\tilde{d} < d$.

Methodological challenges

- Surprisingly: finding a coherent notation and setting that are formally correct, especially for categorical regressors/features.
- Solutions for models containing interaction terms.
- Formalizing existing methods, such as *marginal effects & adjusted predictions* and *average predictive comparisons* to allow for proving that they are a special case of the current framework.

Quantities in the framework

- **generalized marginal effects** - change of conditional expectation, averaged over a certain set of regressor/feature values.
- **individualized expectation** - conditional expectation, averaged over a certain set of regressor/feature values.
- **individualized predictive distribution** - distribution of the target variable with a specific individualized expectation set as expected value.

Advantages compared to existing effect size estimates and visualizations

1. For each definition, the user can individually determine probability measures that represent the situationally appropriate averaging over the inputs.

$\longrightarrow$ Quantities may be seen as *tool kits* that can be specified to derive effects/expectations etc. over areas of interest.

2. After specification, one gets functions of the parameter vector θ for each quantity.

$\longrightarrow$ This allows us to provide a consistent method of calculating point estimates and well-interpretable uncertainty regions.

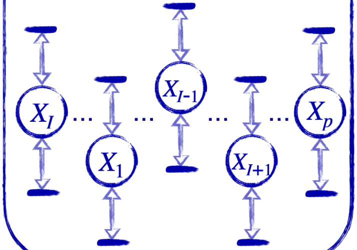
Assumptions

In our framework, a regressor/feature whose effect on the target variable is of interest (X_{I}) has to be chosen for each quantity.

To aide in the choices of probability measures, we have identified the following "assumptions":

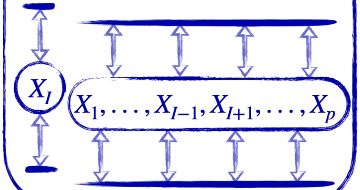
Assumption (A.I) :

All regressors are independent variables.



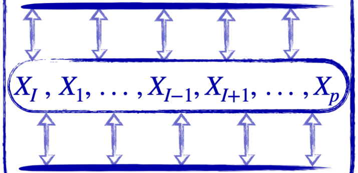
Assumption (A.II') :

While the regressors follow a joint distribution, that of interest is treated as independent.



Assumption (A.II'') :

The regressors follow a joint distribution and are treated as such.



Choices of probability measures

One of the main contributions of the proposed framework is its flexibility to be tailored to most research settings, so there are no universally best choices of probability measure.

Still, here are some examples that will often be reasonably applicable:

- The distribution given by the **relative frequency** in a given data set.
- The (discrete) **Uniform distribution** - for metric regressors this choice additionally has nice computational properties.
- The distribution of characteristics in the population as **determined by previous studies**.

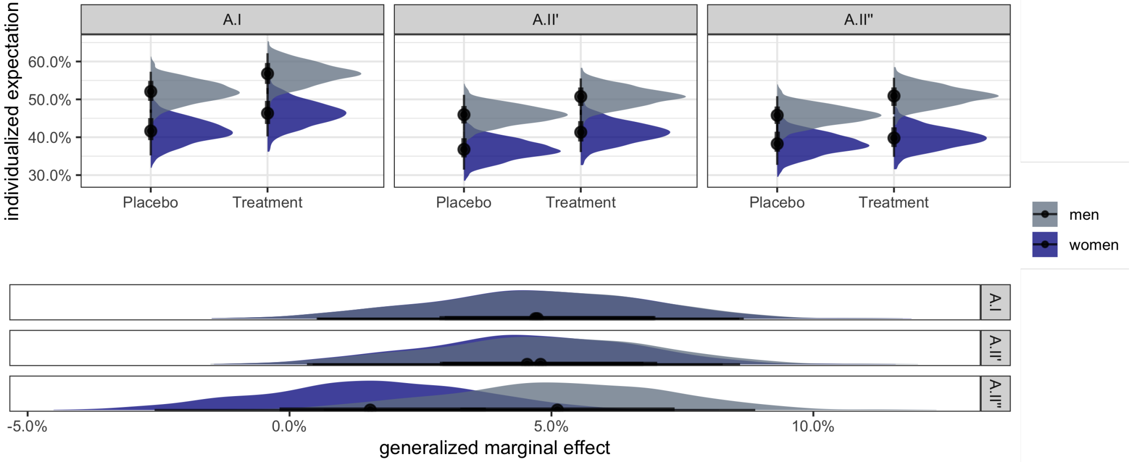
$\Rightarrow$ Each combination of assumption and marginal probability measure choice leads to a different interpretation of the resulting quantity!

Uncertainty modelling

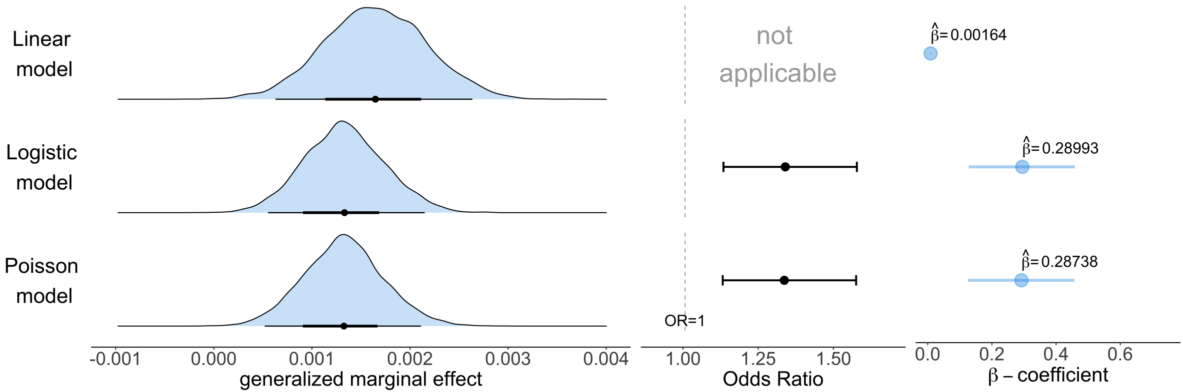
- We take a Bayesian approach to quantifying uncertainty – but all methods can still be applied to the results of both Bayesian and frequentist analyses!
- Specifically, this is achieved by treating θ as a random variable with distribution equal to
    - the posterior distribution given by a Bayesian analysis
    - $N(\hat{\theta}, \Sigma_{\hat{\theta}})$ for the frequentist point estimate $\hat{\theta}$ and covariance matrix $\Sigma_{\hat{\theta}}$

  and then deriving a point estimate (e.g. *median*) and credible set (e.g. *equal-tailed interval*) directly for the random variable defined by $q \circ \theta$, for any considered quantity q.

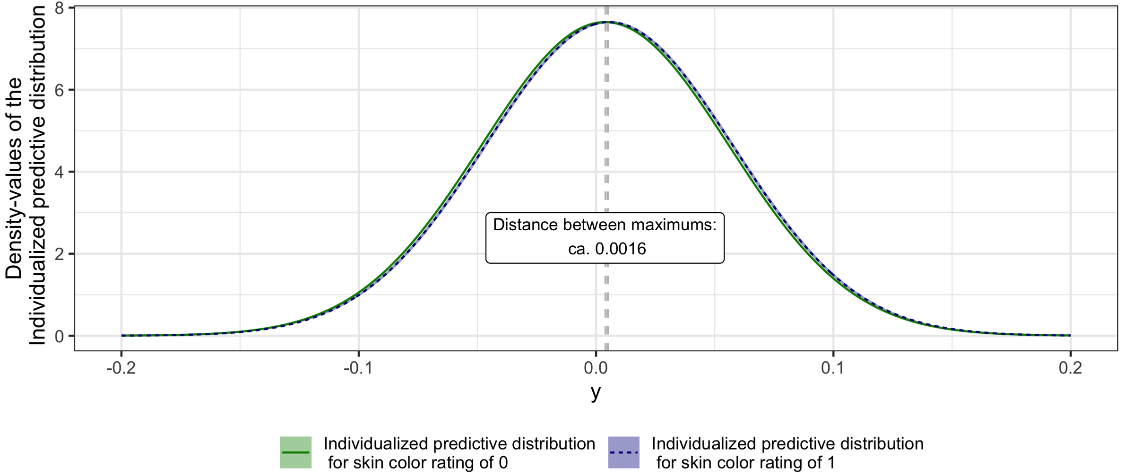
Example: Evidence-based decision making



Example: Multi-Analyst Studies



Example: Visualization through individualized predictive distribution



Outlook

- We are currently working on extending the framework to non- and semi-parametric settings.
- The developed framework can only be meaningful if it is communicated and recommended to practitioners on a large scale. To this end, both user-friendly implementation and tutorial papers aimed at various disciplines are important tools.
- Time-to-event modelling (survival analysis) is an interesting example of settings where the conditional expectation of the target variable is not always readily available – we continue to investigate how said expectation may still be derived in these settings so that they may be incorporated into our framework.

References

Busenbark, John R. et al. (2022). "A Marginal Effects Approach to Interpreting Main Effects and Moderation". In: *Organizational Research Methods* 25.1, pp. 147–169. DOI: 10.1177/1094428120976838. eprint: https://doi.org/10.1177/1094428120976838. URL: https://doi.org/10.1177/1094428120976838.

Frias-Navarro, Dolores et al. (2020). "Replication crisis or an opportunity to improve scientific production?" In: *European Journal of Education* 55.4, pp. 618–631. DOI: https://doi.org/10.1111/ejed.12417.

Gelman, Andrew and Iain Pardoe (2007). "2. Average Predictive Comparisons for Models with Nonlinearity, Interactions, and Variance Components". In: *Sociological Methodology* 37.1, pp. 23–51. DOI: 10.1111/j.1467-9531.2007.00181.x.

Goldstein, Alex et al. (2015). "Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation". In: *Journal of Computational and Graphical Statistics* 24.1, pp. 44–65. DOI: 10.1080/10618600.2014.907095.

Halsey, Lewis G. (2019). "The reign of the *p*-value is over: what alternative analyses could we employ to fill the power vacuum?" In: *Biology Letters* 15.5, p. 20190174. DOI: 10.1098/rsbl.2019.0174.

Kümpel and Hoffmann (2022). "A formal framework for generalized reporting methods in parametric settings". In: arXiv: 2211.02621 [stat.ME].

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin (2016). ""why should i trust you?": Explaining the predictions of any classifier". In: *KDD*.

Roelfs, David J et al. (Jan. 2013). "Meta-Analysis for Sociology - A Measure-Driven Approach.". eng. In: *Bull Methodol Sociol* 117.1, pp. 75–92. ISSN: 0759-1063 (Print); 0759-1063 (Linking). DOI: 10.1177/0759106312465554.

Sandercock, P et al. (2016). "The third International Stroke Trial (IST-3), 2000-2015 [dataset]". In: *University of Edinburgh & Edinburgh Clinical Trials Unit*. URL: https://doi.org/10.7488/ds/1350.

Silberzahn, R. et al. (2018). "Many analysts, one data set: Making transparent how variations in analytic choices affect results". In: *Advances in Methods and Practices in Psychological Science* 1, pp. 337–356. DOI: https://doi.org/10.1177/2515245917747646.

Williams, R. (2012). "Using the margins command to estimate and interpret adjusted predictions and marginal effects". In: *Stata Journal* 12.2, 308–331(24).

OTTAWA 2023

64TH WORLD STATISTICS CONGRESS

Thank you for your attention!