UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DIPARTIMENTO
DI SCIENZE
STATISTICHE

Robust Confidence Intervals for Generalized Linear Models with an Application to RNA-Sequencing Data

---

Andrea Panarotto[1]
✉ `andrea.panarotto@unipd.it`

Joint work with: Riccardo De Santis[1], Livio Finos[1]

[1]Department of Statistical Sciences, University of Padova, Padova, Italy

Statistical Methods for Post Genomic Data – Grenoble, January 30th, 2026

Introduction

- **Confidence intervals** provide essential inferential summaries beyond point estimates in **generalized linear models** (GLMs).
- Interval construction in **parametric modeling** depends on *correct model specification*. Violations of the assumptions may lead to serious *undercoverage* of nominal confidence levels.
- In high-throughput genomic data, *overdispersion*, *heteroskedasticity*, and *unobserved sources of variability* are often on the agenda.
- We propose a **robust approach** for constructing confidence intervals for regression parameters in GLMs.

Setup

Consider a **Generalized Linear Model**

$$g\left(\mathbb{E}\left[Y_i\right]\right) = x_i\beta + \mathbf{z_i}^\top\gamma, \quad i = 1, \ldots, n\,,$$

where

- $y_1, \ldots, y_n$ are independent observations from an exponential dispersion family

$$f\left(y_i; \theta_i, \phi_i\right) = \exp\left\{\frac{y_i\theta_i - b\left(\theta_i\right)}{\phi_i} + c\left(y_i, \phi_i\right)\right\}, \quad i = 1, \ldots, n\,.$$

- x_i *predictor of interest*, $\beta \in \mathbb{R}$;
- $\mathbf{z}_i$ p-dimensional vector of other observed covariates, $\gamma \in \mathbb{R}^p$ (*nuisance parameters*).
- **Goal**: provide a reliable *confidence interval* for β.

Sign-flip score tests

A $(1-\alpha)$-**confidence interval** for a parameter β is the set

$$C_{1-\alpha}(X) = \{\beta_0 \mid p\text{-value}(X; H_0 : \beta = \beta_0) \geq \alpha\} \ ,$$

obtained by *inversion of a α-level test*.

- **Flip-score tests** use as test statistic the *observed score* under the null hypothesis

  $$s_\beta^{obs} = \sum_{i=1}^{n} \nu_i, \quad \nu_i = \frac{\partial}{\partial \beta} \ell\left(\beta, \gamma; y_i\right)\Big|_{\hat{\gamma},\ \beta=\beta_0}$$

  and compares it with the scores with *randomly sign-flipped contributions*

  $$s_\beta^b = \sum_{i=1}^{n} \pm \nu_i \ , \quad b = 2, \ldots, B \, .$$

→ The test that rejects H_0 when $s_\beta^{obs} > s_\beta^{\lceil (1-\alpha)B \rceil}$ is an *asymptotically valid* α-level test.

Robust sign-flip score tests

- Accounting for nuisance estimation ($\hat{\gamma}$) → **effective flipscores** (Hemerik et al., 2020)

  $$S = s_\beta - J_{\beta,\boldsymbol{\gamma}} J_{\boldsymbol{\gamma},\boldsymbol{\gamma}}^{-1} \mathbf{s}_{\boldsymbol{\gamma}} \; ,$$

  where

  - $s_\beta = \frac{\partial}{\partial \beta} \ell(\beta, \boldsymbol{\gamma})$
  - $\mathbf{s}_{\boldsymbol{\gamma}} = \frac{\partial}{\partial \boldsymbol{\gamma}} \ell(\beta, \boldsymbol{\gamma})$
  - $J_{\beta,\boldsymbol{\gamma}}, J_{\boldsymbol{\gamma},\boldsymbol{\gamma}}$ are elements of the Fisher information matrix

- Conservativeness for small sample sizes → **standardized flipscores** (De Santis et al., 2025)

→ The tests are *robust against variance misspecification*.

Inversion of sign-flip score tests

Our construction is based on the inversion of sign-flipping score tests.

- Fix a confidence level $1-\alpha$. Once the estimate $\hat{\beta}_{\text{obs}}$ is obtained, consider the **one-sided tests**

  $$H_0 : \ \beta = \beta_0 \quad \text{vs} \quad H_1 : \ \beta > \beta_0 \qquad \text{for } \beta_0 < \hat{\beta}_{\text{obs}} \,, \tag{1}$$

  and

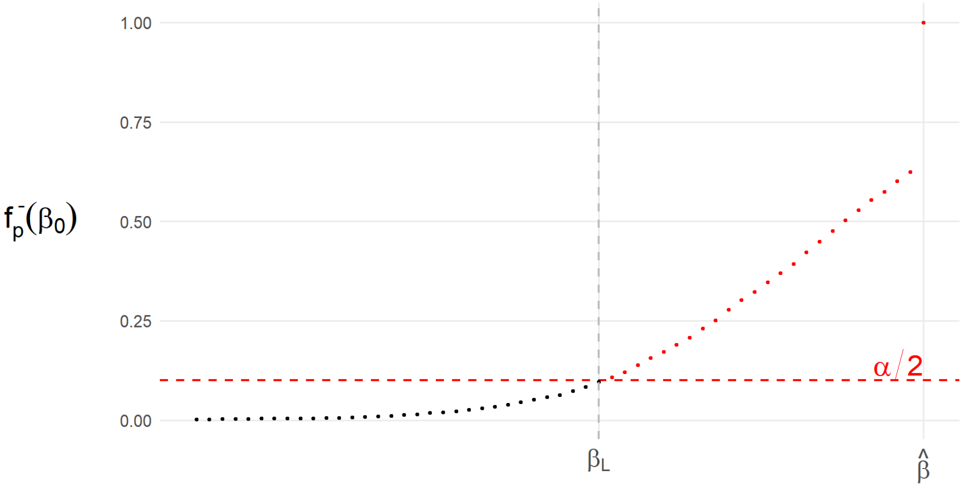
  $$H_0 : \ \beta = \beta_0 \quad \text{vs} \quad H_1 : \ \beta < \beta_0 \qquad \text{for } \beta_0 > \hat{\beta}_{\text{obs}} \,, \tag{2}$$

  at *level* $\alpha/2$.
- Let $f_p^-(\beta_0)$ and $f_p^+(\beta_0)$ be the functions that return the p-value for the test (1), (2), respectively. Then, the confidence bounds for the confidence interval are

$$\beta_L = \inf\left\{\beta_0 < \hat{\beta} : f_p^-(\beta_0) \geq \alpha/2\right\} \,, \qquad \beta_U = \sup\left\{\beta_0 > \hat{\beta} : f_p^+(\beta_0) \geq \alpha/2\right\} \,.$$

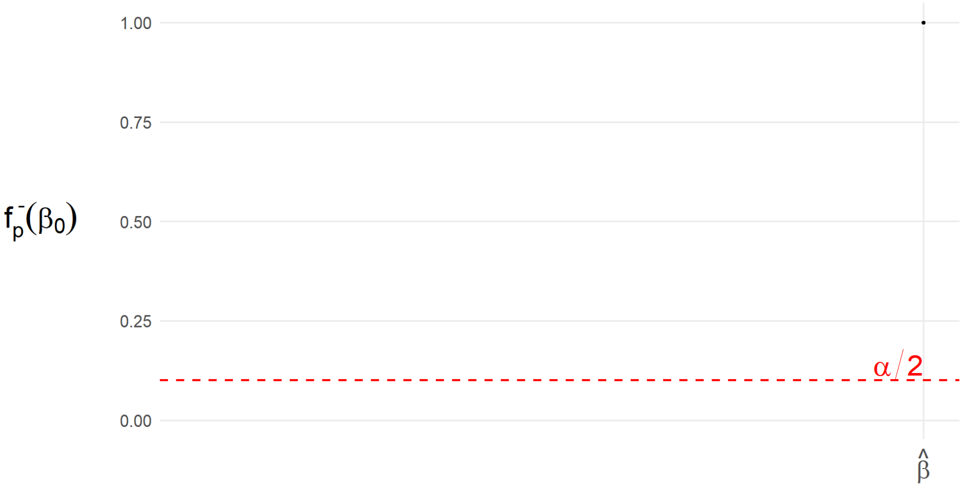
Inversion of sign-flip score tests

$$\beta_L = \inf\left\{\beta_0 < \hat{\beta} : f_p^-(\beta_0) \geq \alpha/2\right\}$$



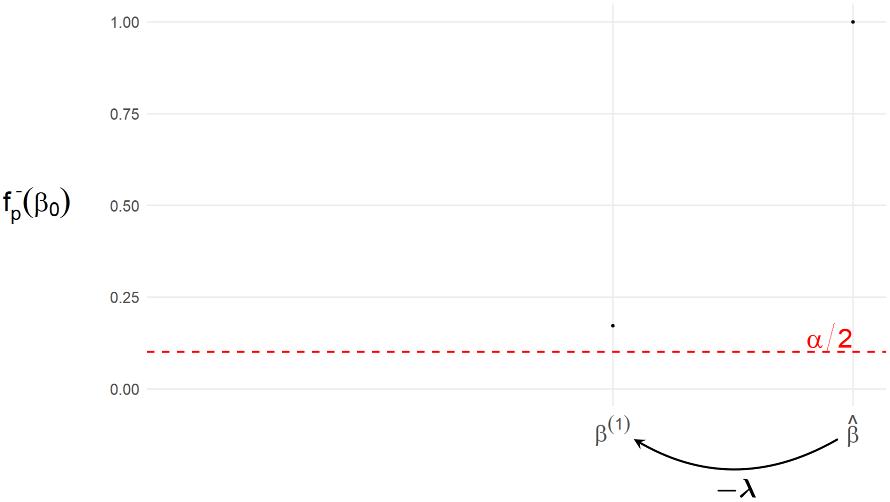
Inversion of sign-flip score tests

$$\beta_L = \inf\left\{\beta_0 < \hat{\beta} : f_p^-(\beta_0) \geq \alpha/2\right\} = \inf\left\{\textbf{?}\right\}$$



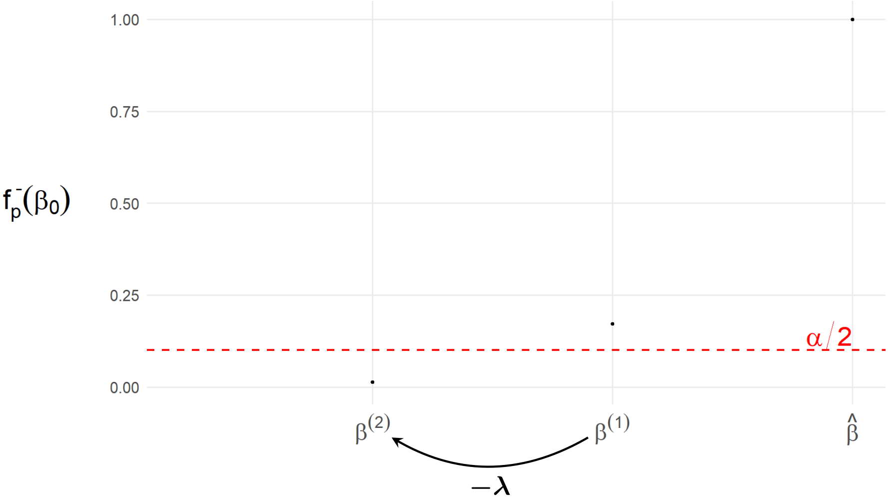
Bisection method

The bounds are obtained repeating the test as the null hypothesis value varies (**same flips!**).



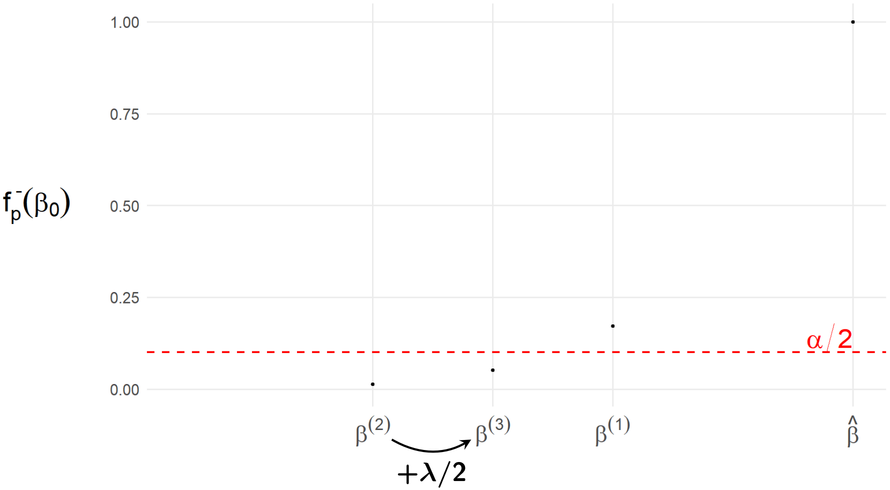
Bisection method

The bounds are obtained repeating the test as the null hypothesis value varies (**same flips!**).



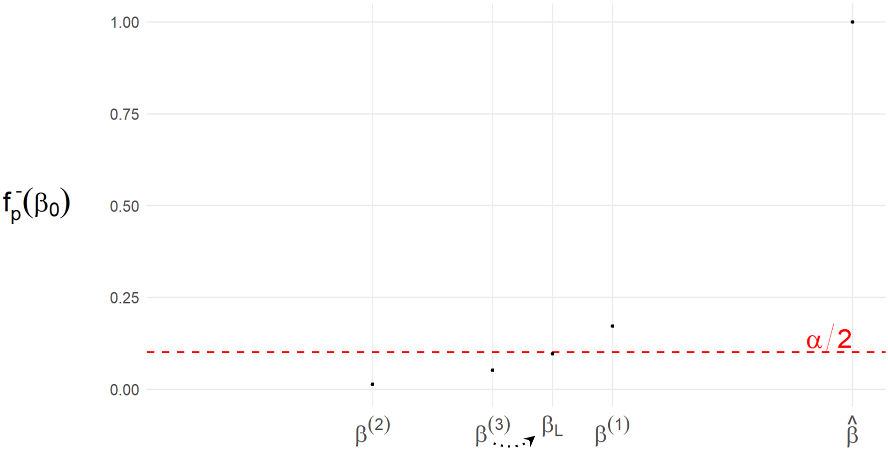
Bisection method

The bounds are obtained repeating the test as the null hypothesis value varies (**same flips!**).

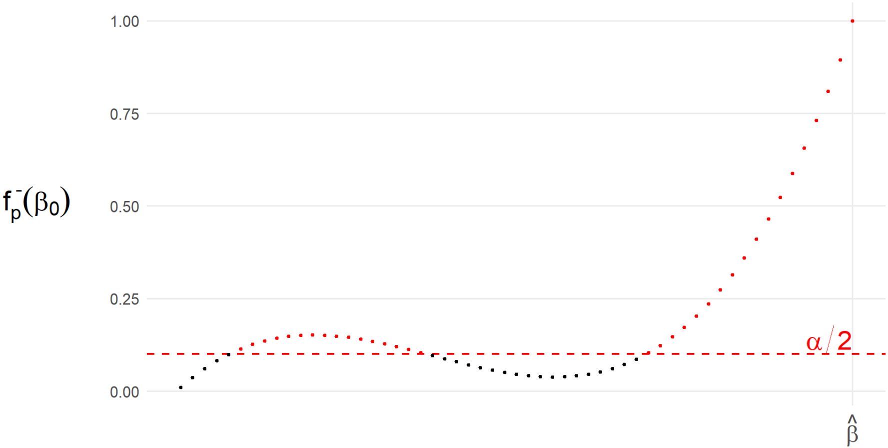


Bisection method

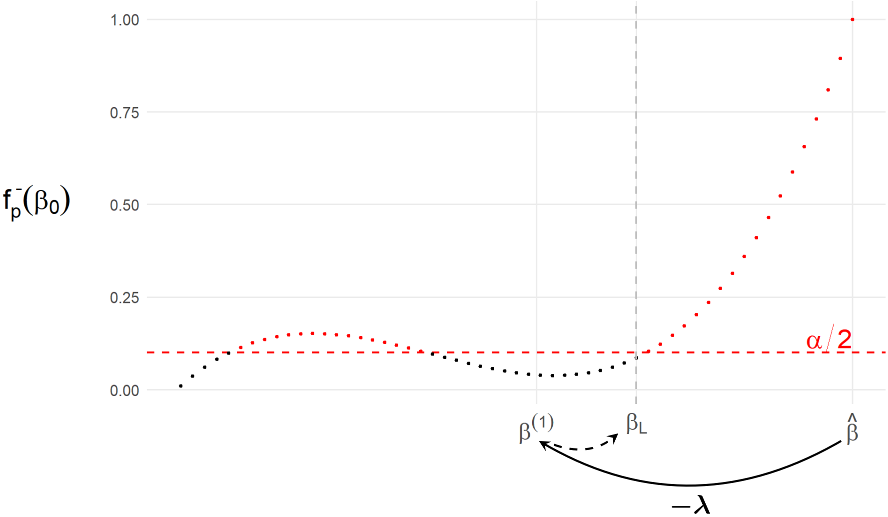
The bounds are obtained repeating the test as the null hypothesis value varies (**same flips!**).



Monotonicity issue



Monotonicity issue



Monotonicity theorems

Theorem (Monotonicity for linear models with effective flipscores test)
Assume that our data are independent and generated by the **linear model**

$$Y_i = x_i\beta + \mathbf{z}_i^T\boldsymbol{\gamma} + \varepsilon_i\,, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2)\,,$$

for $i = 1, \ldots, n$. *Let* $\hat{\beta}_{obs}$ *be the maximum likelihood estimate of* β *and let* $\beta_1 < \beta_0 < \hat{\beta}_{obs}$. *Let* p_k, $k = 0, 1$, *be the p-values associated with the* **effective** *flipscores tests* $H_0 : \beta = \beta_k$ *versus* $H_1 : \beta > \beta_k$, *for* $k = 0, 1$, *respectively, assuming that the tests share the same set of flip matrices. Then,* $p_1 \leq p_0$.

$\rightarrow$ f_p^- and f_p^+ are monotone!

Monotonicity theorems

Corollary (Asymptotic monotonicity for generalized linear models)

Assume that our data are independent and generated by the **generalized linear model**

$$g\left(\mathbb{E}\left[Y_i\right]\right) = x_i\beta + \mathbf{z_i}^\top\gamma, \quad f\left(y_i;\theta_i,\phi_i\right) = \exp\left\{\frac{y_i\theta_i - b\left(\theta_i\right)}{\phi_i} + c\left(y_i,\phi_i\right)\right\}, \quad i = 1,\ldots,n\,.$$

Let $\hat{\beta}$ be the estimate of β and let $\beta_1 < \beta_0 < \hat{\beta}$. Let p_k, $k = 0, 1$, be the p-values associated with the **effective or standardized** *flipscores tests $H_0 : \beta = \beta_k$ versus $H_1 : \beta > \beta_k$, for $k = 0, 1$, respectively. Then, as $n \to \infty$, we definitely have $p_1 \leq p_0$.*

→ f_p^- and f_p^+ are asymptotically monotone!

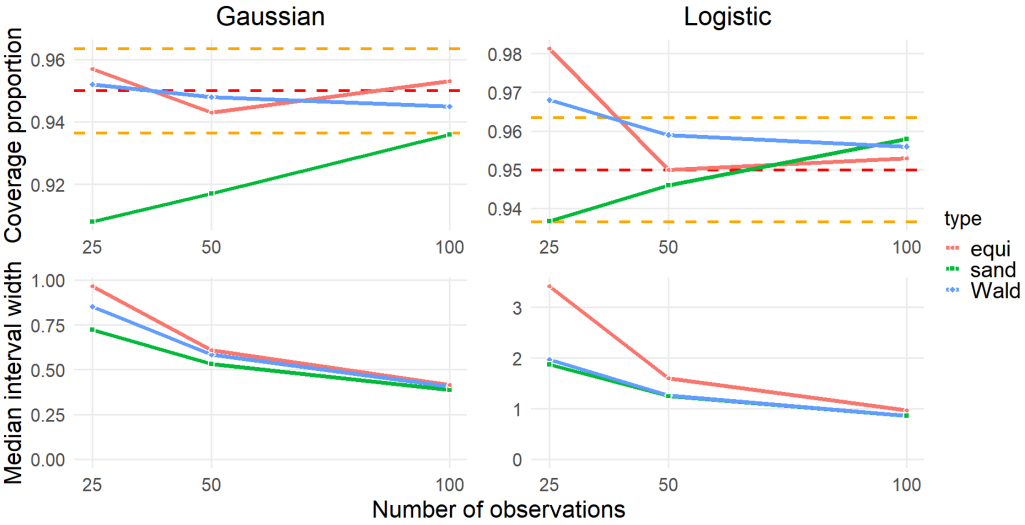
Simulation experiments

We explore six simulation settings.

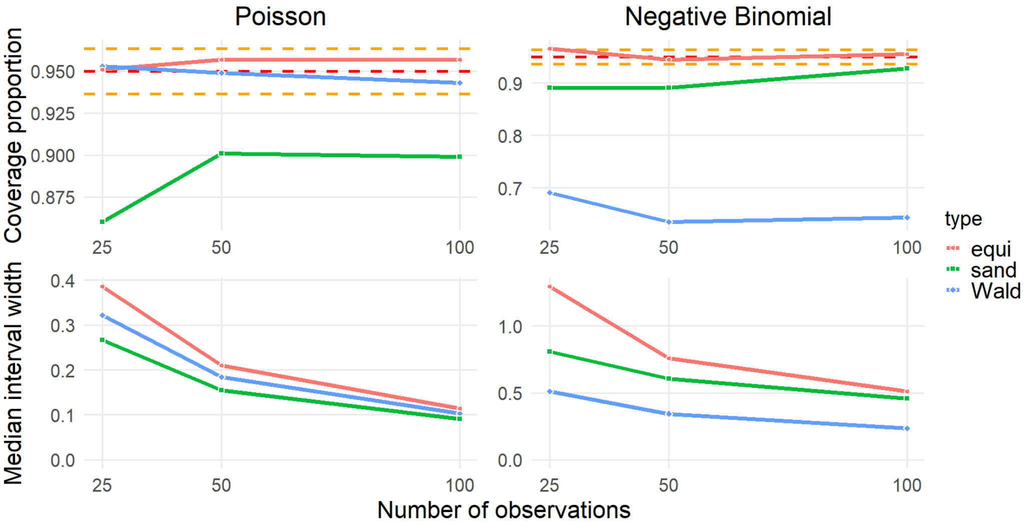
- Correct linear model
- Correct Poisson model
- Linear model with target heteroskedasticity
- Correct logistic model
- Overdispersed Poisson model
- Linear model with nuisance heteroskedasticity

For each case, we repeat 1000 experiments, varying N between 25, 50, and 100. The confidence level is $1 - \alpha = 0.95$.

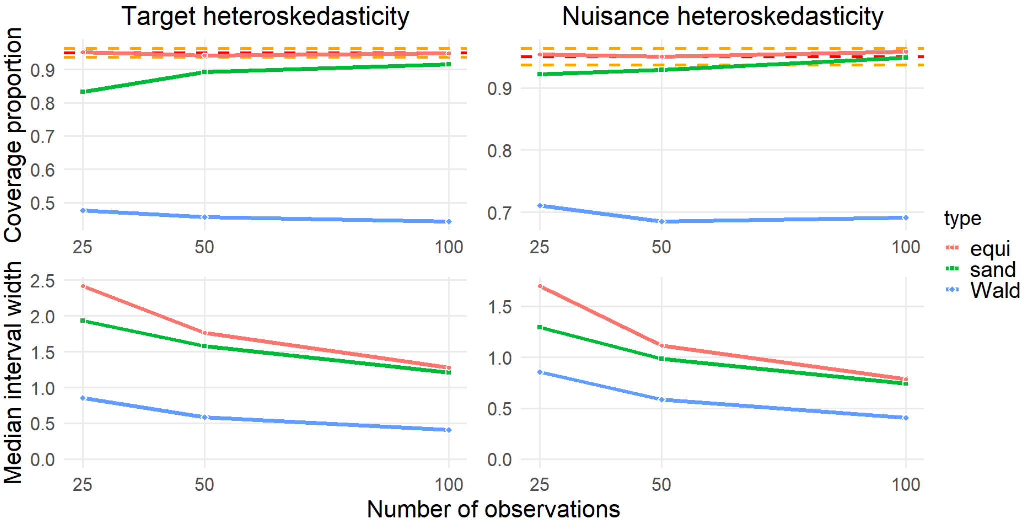
Simulation experiments



Simulation experiments



Simulation experiments



Application

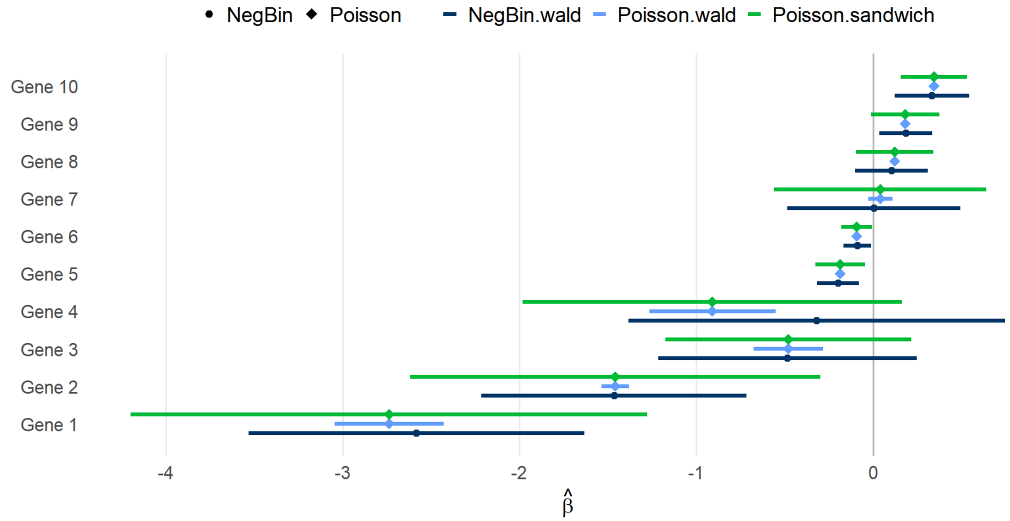
Goal: find differentially expressed genes across a group of units.

Dataset: Liver Hepatocellular Carcinoma from the Cancer Genome Atlas (Tomczak et al., 2015; Erickson et al., 2016) (20000+ genes, 344 patients)

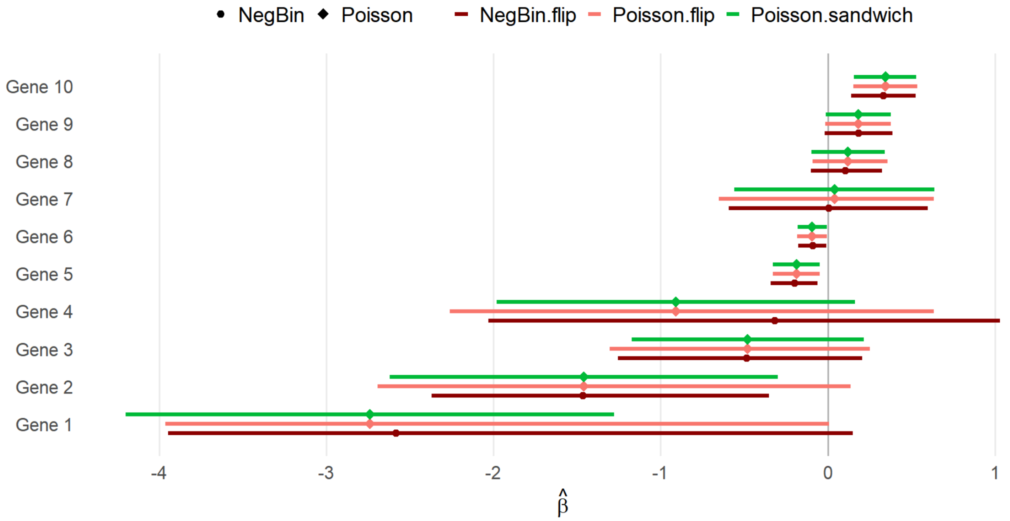
Model: $Y \sim \text{gender} + \text{age} + X$, where Y is the *gene expression* and X indicates whether the tumor is in the *first pathological stage* or in a higher stage. We fit both Poisson and Negative Binomial models.

We compare our intervals to **Wald-type** confidence intervals, built using the classical estimate of the covariance and the (more robust with large n) **sandwich** estimate.

Application



Application



Discussion and conclusion

Summary

- We propose an approach for constructing confidence intervals based on the inversion of sign-flipping hypothesis tests.
- The method guarantees *asymptotic nominal coverage and finite-sample accuracy* while remaining valid under variance misspecification.
- This distributionally robust approach is particularly suited to the *complex variance structures* of modern post-genomic datasets.
- The cost is a *larger computational expense*.

Current and future work

- An alternative is provided for the construction of *symmetric* confidence intervals.
- An extension to *multiple testing* will be addressed in the very near future.

References

Agresti, A. (2015). *Foundations of linear and generalized linear models*. John Wiley & Sons.

Casella, G., McCulloch, C. E., et al. (1984). Confidence intervals for discrete distributions. Technical Report BU-839-M, Cornell University.

De Santis, R., Goeman, J. J., Hemerik, J., Davenport, S., and Finos, L. (2025). Inference in generalized linear models with robustness to misspecified variances. *Journal of the American Statistical Association*, 120(552):2762–2771.

Erickson, B. J., Kirk, S., Lee, Y., Bathe, O., Kearns, M., Gerdes, C., Rieger-Christ, K., and Lemmerman, J. (2016). The cancer genome atlas liver hepatocellular carcinoma collection (tcga-lihc) (version 5) [data set]. The Cancer Imaging Archive.

Hemerik, J. and Goeman, J. J. (2018). Exact testing with random permutations. *Test*, 27(4):811–825.

Hemerik, J., Goeman, J. J., and Finos, L. (2020). Robust testing in generalized linear models by sign flipping score contributions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(3):841–864.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Salvan, A., Sartori, N., and Pace, L. (2020). *Modelli Lineari Generalizzati*, volume 124 of *UNITEXT*. Springer Milan, Milano.

Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). Review The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Współczesna Onkologia*, 1A:68–77.

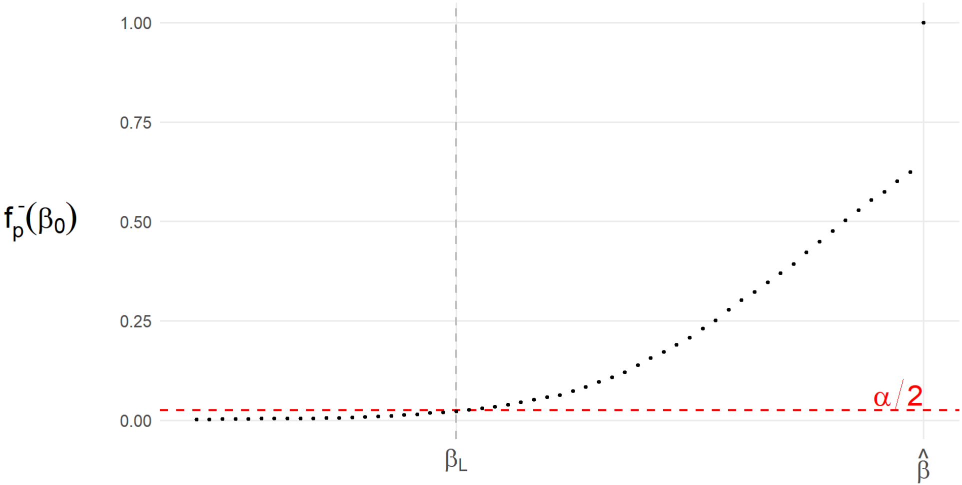
Thank you for your attention.

The method will be soon available in the `flipscores` package, on CRAN and on `https://github.com/livioivil/flipscores`.

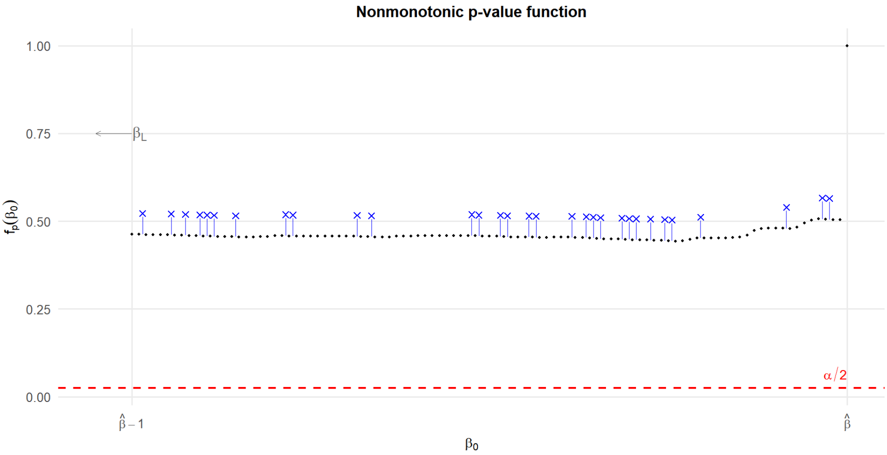
Backup slides

Inversion of sign-flip score tests

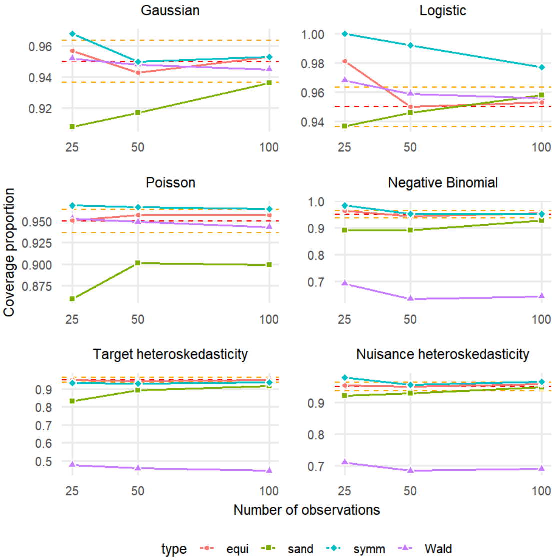
$$\beta_L = \inf\left\{\beta_0 < \hat{\beta} : f_p^-(\beta_0) \geq \alpha/2\right\}$$



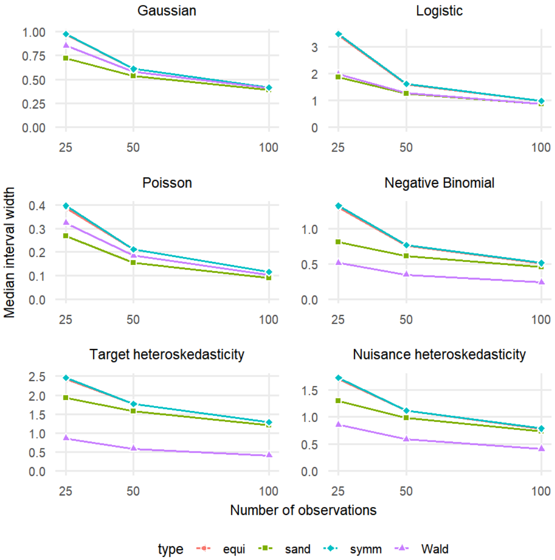
Monotonicity issue



Symmetric


Gaussian
Logistic
Poisson
Negative Binomial
Target heteroskedasticity
Nuisance heteroskedasticity
Coverage proportion
Number of observations
25
50
100
0.96
0.94
0.92
1.00
0.98
0.96
0.94
0.950
0.925
0.900
0.875
1.0
0.9
0.8
0.7
0.6
0.5
type
equi
sand
symm
Wald

Symmetric


Gaussian
Logistic
Poisson
Negative Binomial
Target heteroskedasticity
Nuisance heteroskedasticity
Median interval width
Number of observations
1.00
0.75
0.50
0.25
0.00
0
1
2
3
0.4
0.3
0.2
0.1
0.0
1.0
0.5
0.0
2.5
2.0
1.5
1.0
0.5
0.0
1.5
1.0
0.5
0.0
25
50
100
type
equi
sand
symm
Wald