Robust testing in generalized linear models by sign-flipping score contributions

Livio Finos
University of Padova

Outline

1. Motivation
2. Flip score test (known nuisance parameters)
3. Effective Flip score test
4. Standardized Flip Score Test
5. Extensions (Multivariate + Clustered)

Permutation tests

Usually

- require less assumptions
- exact control of Type I Error, even for small sample size
- converge to parametric counterpart
  (ie asymptotically same power)
- multivariate (multiplicity correction): easy and powerful

A major limitation: Continuous Confounders

Example: Poisson model ($g() = log()$ link function)

$$log(E(y_i)) = g(\mu) = \eta = \gamma_0 + \gamma_1 z_i + \beta x_i, \;\; i = 1, \dots, n$$

$$H_0: \;\beta = 0 \;\; \forall\, \gamma = (\gamma_0, \gamma_1)$$

- Z is **categorical**: y_i are exchangeable **within levels**. Permute y_i within levels to get the null distribution.
- Z is **continuous**: Residualization is problematic: $V(y_i) = e^{\gamma_0 + \gamma_1 z_i}$ depends on the covariates.

Solutions with some limitations: Gail et al., (1988), Heller et al., (2009), Parhat et al. (2014), Pauly et al (2015).

Flip Score test

In this presentation we focus on GLM:

- Effective Flip Score test: Hemerik, Goeman, Finos. *Robust testing in generalized linear models by sign-flipping score contributions*. JRSS-B `doi:10.1111/rssb.12369`
- Standardized Flip Score test: De Santis, Goeman, Hemerik, Finos
  `https://arxiv.org/abs/2209.13918`.

Both approaches are:

- **Robust**: allow for biased variance estimators (i.e. overdispersion, heteroscedastic, etc)
- **Flexible**: general approach, many extensions

Outline

The parametric score test

$$
\begin{aligned}
Score &= \left.\frac{\partial \ell(\beta|x,z,y)}{\partial \beta}\right|_{\beta=0} \\
&= \sum_{i=1}^{n} \frac{\partial}{\partial \beta} \log f_\gamma(y_i) = \\
&= \sum_{i=1}^{n} \nu_i \underset{H_0}{\dot\sim} N(0,\mathcal{I}) \ \ \text{(asymptotically, under } H_0\text{)} \\
&\quad \mathcal{I}\text{: Fisher Information Matrix}
\end{aligned}
$$

$\nu_i \sim NOTnormal(0, var(\nu_i))$ + Centr.Lim.Thm =
Approximated Type I Error control, but very good in practice.

Sign-Flip Score test with known γ_0 and γ_1

$$
\begin{aligned}
S^* : &= \sum_{i=1}^{n} \pm\nu_i \\
(Poisson) &= \sum_{i=1}^{n} \pm x_i (y_i - e^{\gamma_0 + z_i \gamma_1}) = \sum_{i=1}^{n} \pm x_i (y_i - \mu_i)
\end{aligned}
$$

- Observed test statistic: $S^{obs} = \sum_{i=1}^{n} +\nu_i$
- p-value: $\frac{\#(|S^*| \geq |S^{obs}|)}{\#resamplings}$

Properties

Second-moment Null-invariance (NI2)

- 0-mean: $E(y_i - \mu_i) = 0 \Rightarrow E(\pm x_i(y_i - \mu_i)) = 0$
- $y_i \perp\!\!\!\perp y_i$
- constant variance:
  $V(+x_i(y_i - \mu_i)) = V(-x_i(y_i - \mu_i)) \Rightarrow V(S^*) = V(S^{obs})$

Properties

It converges to parametric score test (i.e. asymptotically):

i. is normal $N(0, \mathcal{I})$
ii. is exact
iii. is locally most powerful (LMP)
iv. if the parametric S test is UMP (UMPU),
    S^* is asymptotically UMP (UMPU)

Outline

Flip score test with UNknown nuisance parameters

- Model: $log(E(y_i)) = \gamma_0 + \gamma_1 z_i + \beta x_i, \;\; i = 1, \ldots, n$
- $H_0: \;\beta = 0 \;\forall\; \gamma = (\gamma_0, \gamma_1)$

In most applications, γ is unknown and we **plug-in an estimate** (under H_0).

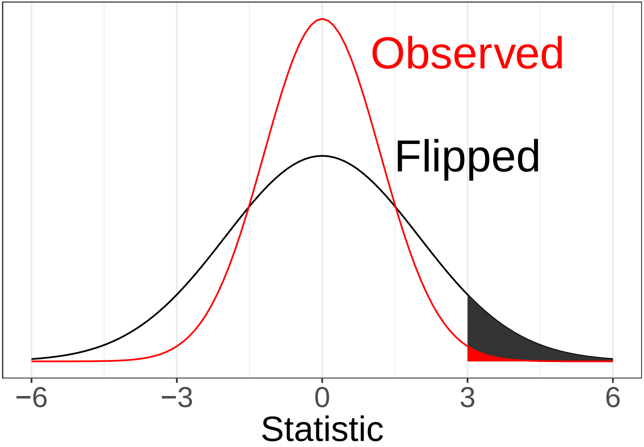
e.g. Poisson model:

$$S = \sum_{i=1}^{n} x_i \left(y_i - e^{\hat{\gamma}_0 + z_i \hat{\gamma}_1} \right) = \sum_{i=1}^{n} x_i \left(y_i - \hat{\mu}_i \right)$$

Does it has any consequences?

YES, Observed test statistic has lower variance than flipped ones :(

The test becomes conservative



Solution: Effective flip score test (or doubly residualized)

YES, Observed test statistic has lower variance than flipped ones :(

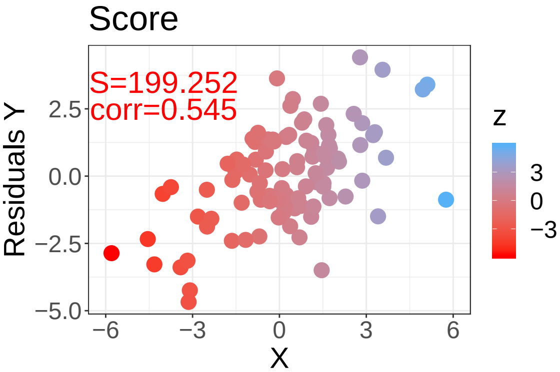
Intuition for **linear model**
$H = Z(Z'Z)^{-1}Z'$

$$
\begin{aligned}
S^*(d(\pm 1)) &= \sum_{i=1}^{n} \pm x_i(y_i - \hat{\mu}_i) \\
&= X'd(\pm 1)(I-H)Y = \\
&= X'(I-H+H)'d(\pm 1)(I-H)Y = \\
&= \mathbf{X}'(\mathbf{I}-\mathbf{H})\mathbf{d}(\pm\mathbf{1})(\mathbf{I}-\mathbf{H})\mathbf{Y} + X'Hd(\pm 1)(I-H)Y
\end{aligned}
$$

Observed score test: $X'Hd(+1)(I-H)Y = X'H(I-H)Y = 0$
Random flip: $X'Hd(\pm 1)(I-H)Y \neq 0$
(0 mean, positive variance)

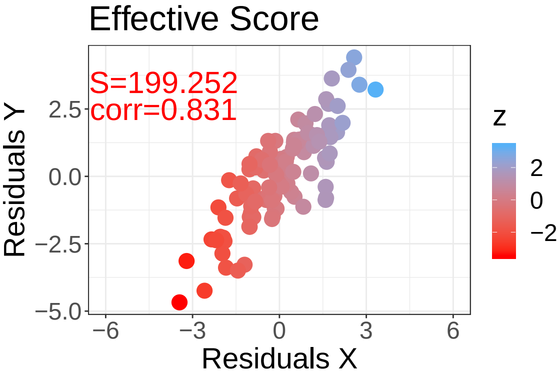
Toy Data

Linear model $y \sim 1 + z + x + \varepsilon, \quad \varepsilon \sim N(0, 1)$, $\mathbf{corr}(\mathbf{x}, \mathbf{z}) = \mathbf{.8}$



Toy Data

Linear model $y \sim 1 + z + x + \varepsilon, \quad \varepsilon \sim N(0,1)$, $\mathbf{corr}(\mathbf{x}, \mathbf{z}) = .8$



Double residualization in GLM

- W diagonal matrix with $w_i = (\partial\mu_i/\partial\eta_i)^2/var(y_i)$
- $I - H = I - WZ(Z'WZ)^{-1}Z'$

The Score test statistic:

$$S^{obs} = X'(Y - \hat{\mu})$$

The **Effective** (Observed) Score test statistic:

$$S_E^{obs} = X'(I - H)(Y - \hat{\mu})$$

and Flipped: $S_E^* = X'(I - H)d(\pm 1)(Y - \hat{\mu})$

General Solution: Effective Score

For every $1 \leq i \leq n$, let

$$\nu_{\hat{\gamma},i}^{(k-1)} = \frac{\partial}{\partial \gamma} \log f_{\beta,\gamma,X_i}(Y_i)\Big|_{\beta=0,\gamma=\hat{\gamma}} \in \mathbb{R}^{k-1},$$

Then

$$S^*_{\hat{\gamma}} = \sum_i \pm(\nu_{\hat{\gamma},i} - \hat{\mathcal{I}}'_{XZ}\hat{\mathcal{I}}_{ZZ}^{-1}\nu_{\hat{\gamma},i}^{(k-1)})$$

with $\hat{\mathcal{I}}$ the Observed Fisher Information under H_0:

$$\hat{\mathcal{I}} = \begin{bmatrix} \hat{\mathcal{I}}_{XX} & \hat{\mathcal{I}}'_{XZ} \\ \hat{\mathcal{I}}_{XZ} & \hat{\mathcal{I}}_{ZZ}. \end{bmatrix}$$

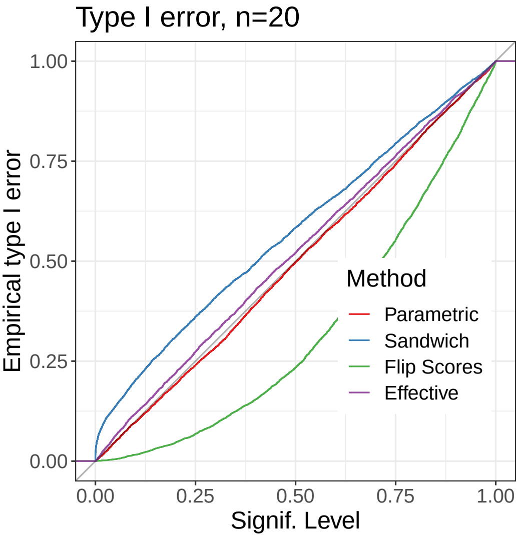
Remark: The effective score is strongly connected with the (parametric) Efficient Score (Cox and Hinkley, 1979).

Simulation: Poisson

Setting

- $y_i \sim Poisson(\mu_i = \exp(1 + 1 \cdot z_i + \beta x_i))$,
- $H_0 : \beta = 0 \ \forall(\gamma_0, \gamma_1)$
- Simulation under $H_0 : \beta = 0$ (nuisance $\gamma_1 = \gamma_0 = 1$)
- (x, z) multivariate normal with high correlation: $corr(xz) = 0.8$ (and variance 1)
- 2000 sign-flips, 10000 datasets

Poisson



Type I error, n=20
Empirical type I error
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
Signif. Level
Method
Parametric
Sandwich
Flip Scores
Effective

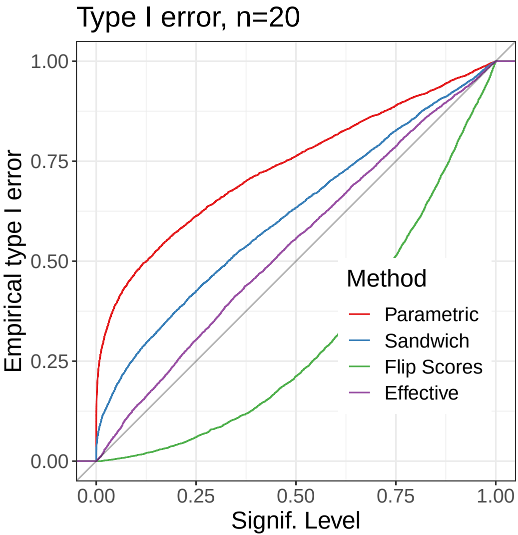
Simulation: overdispersion

The Effective Score Statistic is robust to biased estimate of W (i.e. variance).

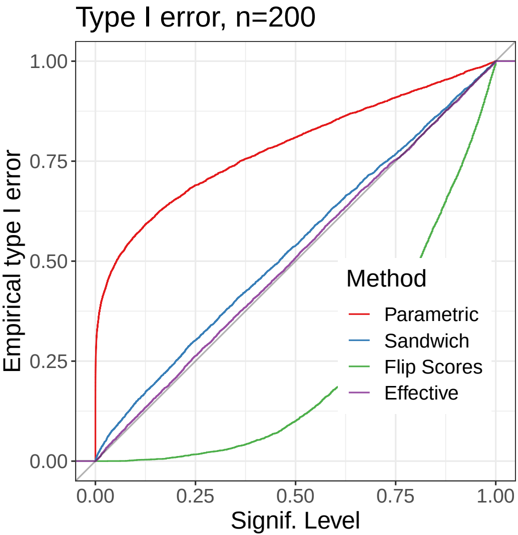
Setting

- $y_i \sim NegBinom(\mu_i = \exp(1 + 1 \cdot z_i + \beta x_i), \phi = 1.2)$,
- $H_0: \ \beta = 0 \ \forall(\gamma_0, \gamma_1)$ but WRONG Poisson model used
- Simulation under $H_0 : \beta = 0$ (nuisance $\gamma_1 = \gamma_0 = 1$)
- (x, z) multivariate normal with high correlation: $corr(xz) = 0.8$ (and variance 1)
- 2000 sign-flips, 10000 datasets

Overdispersion



Overdispersion



Effective Flip Score is Asymptotically Exact

- We need: $\nu_i \perp\!\!\!\perp \nu_{i'}$.
- However, when we plug $\hat{\gamma}$ into $S_{\hat{\gamma}}$, the $\nu_{\hat{\gamma},i}$ become dependent (e.g. in linear model the effective d.f. are $n - \mathrm{rank}(Z)$).
- The correlation disappears when $n \to \infty$

We get **asymptotically exact** test (with any variance estimates).

Effective Flip Score is Asymptotically Exact

- We need: $\nu_i \perp\!\!\!\perp \nu_{i'}$.
- However, when we plug $\hat{\gamma}$ into $S_{\hat{\gamma}}$, the $\nu_{\hat{\gamma},i}$ become dependent (e.g. in linear model the effective d.f. are $n - \mathrm{rank}(Z)$).
- The correlation disappears when $n \to \infty$

We get **asymptotically exact** test (with any variance estimates).

We propose the **Standardized** Flip Score for small (and large) sample size!

Outline

Standardized Flip Scores

It is easy to derive: $Var\{S(d(\pm 1))\} =$
$n^{-1}X^TW^{1/2}(I-H)d(\pm 1)(I-H)d(\pm 1)(I-H)W^{1/2}X + o_p(1)$
Standardized flipscores:

$$S^*(d(\pm 1)) = S(d(\pm 1))/Var\{S(d(\pm 1))\}^{1/2}$$

Results: the anti-conservativeness disappears:

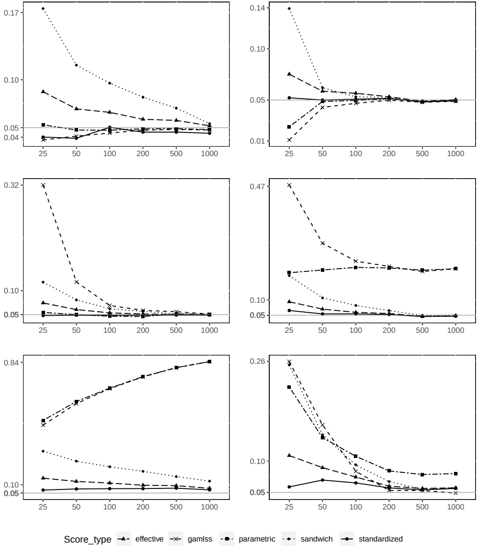
- with unbiased estimate of $\hat{\mu}$: **Second Moment Null Invariant (NI2)** (almost exact, see later)
- with unbiased estimate of $\hat{\mu}$ + normality (lm): **Exact**

Simulation: Type I error control

Top: correct Poisson and Logistic model.

Middle Normal, heteroscedasticity nuisance / target.

Bottom: false Poisson model, two groups Negative-binomial.



Standardized vs Effective Flip Scores

Property: **second-moment null-invariant (NI2)** means that S^* and S^{obs} have zero mean and same variance (i.e. it is not an exact test, same as for parametric score).

When $\hat{W}$ is **unbiased** estimate of true W:

- **Effective** flip Score is only *Asymtotically* NI2
- **Standardized** flip Score is (finite sample) NI2

When $\hat{W}$ is **biased** estimate of true W:

- **Effective** flip Score is only *Asymtotically* NI2
- **Standardized** flip Score is Asymtotically NI2,
  but it converges faster.

Standardized is anyway better than Effective.

Outline

Multivariate

Multivariate Hypotheses

$$H_{0j}:\ \beta_j = 0 \mid (\gamma_{0j}, \tau_{0j}) \in \mathbb{G}_j \times \mathbb{T}_j, \quad H_0 : \bigcap_{j=1}^{q} H_{0j} \tag{1}$$

$$H_{1j}:\ \beta_j > 0 \mid (\gamma_{0j}, \tau_{0j}) \in \mathbb{G}_j \times \mathbb{T}_j, \quad H_1 : \bigcup_{j=1}^{q} H_{1j}$$

Multivariate Test Statistic:

$$S^*_{Oj} = X'_j(I - H)Ud(\pm)U'(Y - \hat{\mu}), \quad S_O = (S_{O1}, \dots, S_{Oq})'$$

Resampled S^*_O: same sign-flipping $d(\pm)$ for all tests $S^*_{O1}, \dots, S^*_{Oq}$.
This ensure same covariance, that is **Multivariate NI2**.

Multivariate: example

- $log(E(Y)) = Z\gamma + X_1\beta_1$
- $log(E(Y)) = Z\gamma + X_2\beta_2$

$$H_0: \ \beta_1 = 0 \cap \ \beta_2 = 0$$

i.e. Same estimate under H_0 (= same H and W)
Examples:

- post-hoc ANOVA with covariates (Z)
- GWAS: case-control (binomial Y) with individual covariates (Z)

Clustered Observations (e.g. mixed-models)

Examples: Mixed-models, Repeated measures, GEE
Clustered observations: true W is BLOCK matrix (not diagonal).
Solution: Flipping matrix $d(\pm 1)$ has same sign within blocks (i.e. signs are flipped block-wise)
Remarks:

- doesn't need balanced designs
- when correlation within clusters is misspecified:
  GEE has unbiased estimate, but loses type I error control
  Flipscores has unbiased estimate AND good type I error control

R package: `flipscores`

Same syntax as `glm()`

An Example:

- Response: Binomial, logit link
- Predictors: x(3-groups) + z(continuous)+interaction
- Clustered observations (15 clusters)

```
set.seed(1)
x=factor(rep(LETTERS[1:3],15))
D=data.frame(y=rbinom(45,1,.05+(x=="C")*.8),x=x,
             z=rnorm(45),id=rep(1:15,each=3))
```

R package: flipscores

```
library(flipscores)
mod=flipscores(y~x*z,data=D,family = binomial, data = D, id = D$id)
summary(mod)
```

```
##
## Call:
## flipscores(formula = y ~ x * z, family = binomial, data = D,
##     id = D$id)
##
## Coefficients:
##                Estimate      Score Std. Error z value Pr(>|z|)
## (Intercept) -2.057e+01 -7.211e+00  2.772e+00  -2.601    0.054 .
## xB          -6.564e-11 -5.673e-16  4.881e-16  -1.162    0.384
## xC           2.195e+01  5.861e+00  3.239e+00   1.810    0.020 *
## z           -4.722e-09 -3.166e-16  3.952e-16  -0.801    0.442
## xB:z         4.722e-09 -5.008e-16  4.076e-16  -1.229    0.278
## xC:z        -5.465e-03  2.433e-16  5.275e-09   0.000    1.000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

R package: flipscores

```
anova(mod,type=3)

## Analysis of Deviance Table: Type III test
## Model: binomial, link: logit
## Inference is provided by FlipScores approach (5000 sign flips).
##
## Response: y
##      Score Df Pr(>Score)
## x   7.6516  2     0.0020 **
## z   0.5828  1     0.4484
## x:z 0.0001  2     0.9904
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion

Standardized Flip Score Test

- for GLM
- gooood control of the type I error

Effective Flip Score Test:

- general: any score statistics
- Asymtotically exact

Both

- assume the link function to be right ($E(\hat{\mu}) = \mu$)
- very **robust** (heteroscedasticity, overdispersion etc)
- Multivariate: easy (even $\mathbf{p} >> \mathbf{n}$)
- quite **fast** to compute
- R package: `https://github.com/livioivil/flipscores`
- some relationship with solution of F. Pesarin (2001) to Behrens-Fisher problem

Extensions

- Spatial/Penalized Models (Ferracioli, Sangalli, Finos. Nonparametric tests for semiparametric regression models DOI, 10.1007/s11749-023-00868-9)
- Confidence Intervals (??)
- more on Multivariate (with A. Vesely; arXiv:2210.02794 + arXiv:2205.12563)
- Median unbiased estiamators (with E.C. Kenne Pagui)
- survival analysis (De Santis), others non-linear models ...