Simulation Description:

Checking performance of Poisson regression

Data Generation (All Settings)

$X \sim \text{Norm}(0,1)$ $\quad \beta = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

$X^* \sim$ one of 4

$Z \sim \text{Norm}(0,1)$ $\quad \lambda = \exp(\beta^T \text{data})$

$Y \sim \text{Pois}(\lambda)$ $\quad$ data is $[1 \times 2]$

seed = 1031

Data Generation (Variations)

15 total

increase n: 100, 1000, 10000

X^* #1: $= X$

X^* #2: $= X + \text{rnorm}(0,1)$

X^* #3: $= X + Z + \text{rnorm}(0,1)$

X^* #4: $= XZ + \text{rnorm}(0,1)$

X^* #5: $= X + \text{rnorm}(\text{mean}(Z), 1)$

Assumptions

X^* is fully observed

model is correctly specified (ie Poisson)

independence of observations

Methods Considered

regular Poisson regression

Estimands

true values

$\beta = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

Starting Values

$\beta = [2,3,4]^T$

What to Save From Each Replication

error setting

n

$\hat{\beta}$ components

$SE(\hat{\beta})$ components

How to Summarize Performance

- histogram of $\hat{\beta}$
- $\hat{\beta}$ mean
- SE mean
- $\hat{\beta}$ sd
- SE sd

What We Expect to See

X^*1: as $n \to \infty$, good → great

X^*2: bad

X^*3-5: worse

} improvement as $n \uparrow$?