Confidence Intervals

Standard vs. Bootstrap

Jose de los Rios, Patrick Swain, Lap Pham, Maxwell Snodgrass

Data Generation

```r
# Generate random data for the Poisson GLM model
generate_data <- function(N) {
# Generate the binary predictor (delta) with 60% probability of being 1
  delta <- rbinom(N, 1, 0.6)

 # Generate x based on chi-squared distributions - 4 df when delta=1, 2 df when delta=0
  x <- ifelse(delta == 1, rchisq(N, df = 4), rchisq(N, df = 2))

# True parameter values
  beta0 <- -1
  beta1 <- -1
  beta2 <- 1/2

# Calculate lambda mean of Poisson
  lambda <- exp(beta0 + beta1 * delta + beta2 * x)

# Generate Poisson outcomes
  y <- rpois(N, lambda)

# Return data frame
  data.frame(y = y, delta = delta, x = x)
}

# Calculate transformation probability y < 2 when delta=1 and x=4
calculate_transformation <- function(beta0, beta1, beta2) {
  lambda <- exp(beta0 + beta1 * 1 + beta2 * 4)
  ppois(1, lambda)  # P(Y ≤ 1) = P(Y = 0) + P(Y = 1)
}
```

We generated datasets that included a binary predictor delta drawn from a Bernoulli distribution with success probability 0.6

We generated a continuous predictor x from a chi-squared distribution with 4 degrees of freedom when delta=1 and 2 degrees of freedom when delta=0

The response variable y is generated from a Poisson distribution with mean λ, defined as

$$\lambda_i = \exp\{\beta_0 + \beta_1 \delta_{i,1} + \beta_2 x_{i,2}\}$$

- $\beta_0 = -1$
- $\beta_1 = -1$
- $\beta_2 = \frac{1}{2}$

Standard

We fitted a Poisson GLM to the generated dataset and computed confidence intervals for each regression coefficient using the normal approximation and the model's estimated standard errors

It also calculates a transformed probability representing a nonlinear function of the GLM parameters, handled via the delta method

Bootstrap

We performed non-parametric bootstrapping with 1,000 resamples to estimate empirical distributions of the model coefficients and the same transformation as the standard inference transformation

For each resample, a Poisson GLM is fit and the relevant parameters and transformation are cached

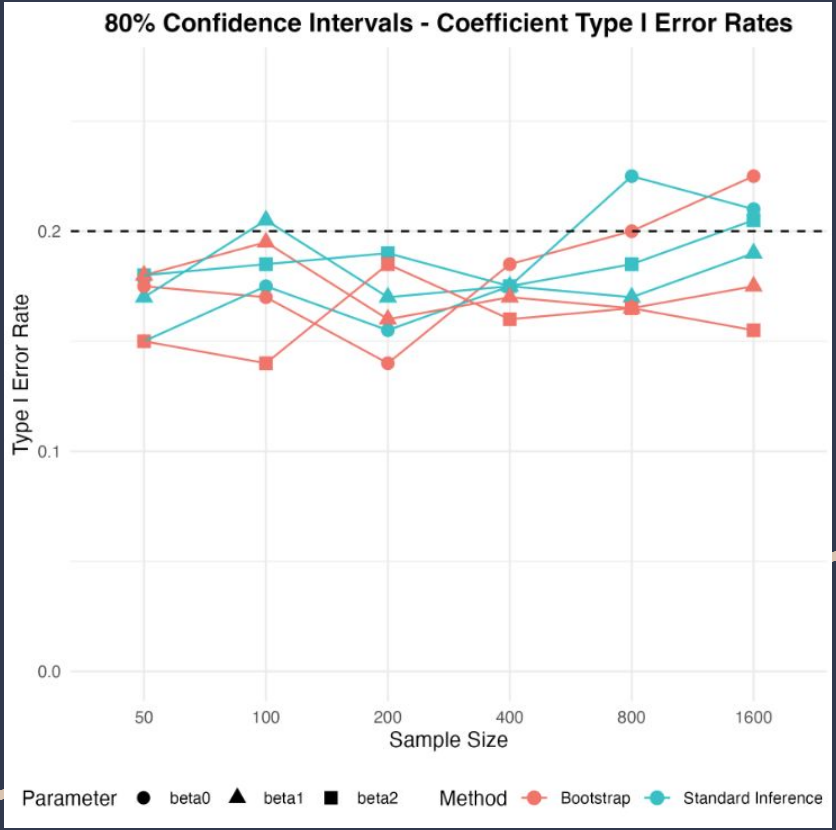
We computed confidence intervals from the bootstrap distribution using the boot package

We repeated these functions over the specified parameters using both methods and stored the results in a .csv before visualizing them

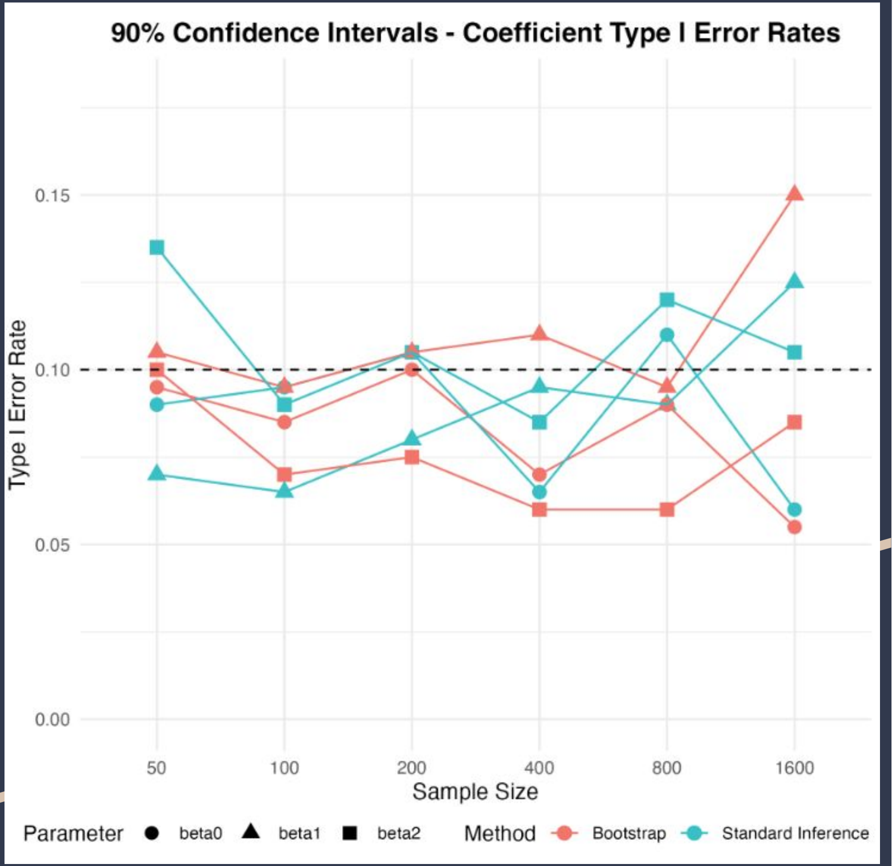
N: [50,100,200,400,800,1600]

Confidence levels: [0.8,0.9,0.95]

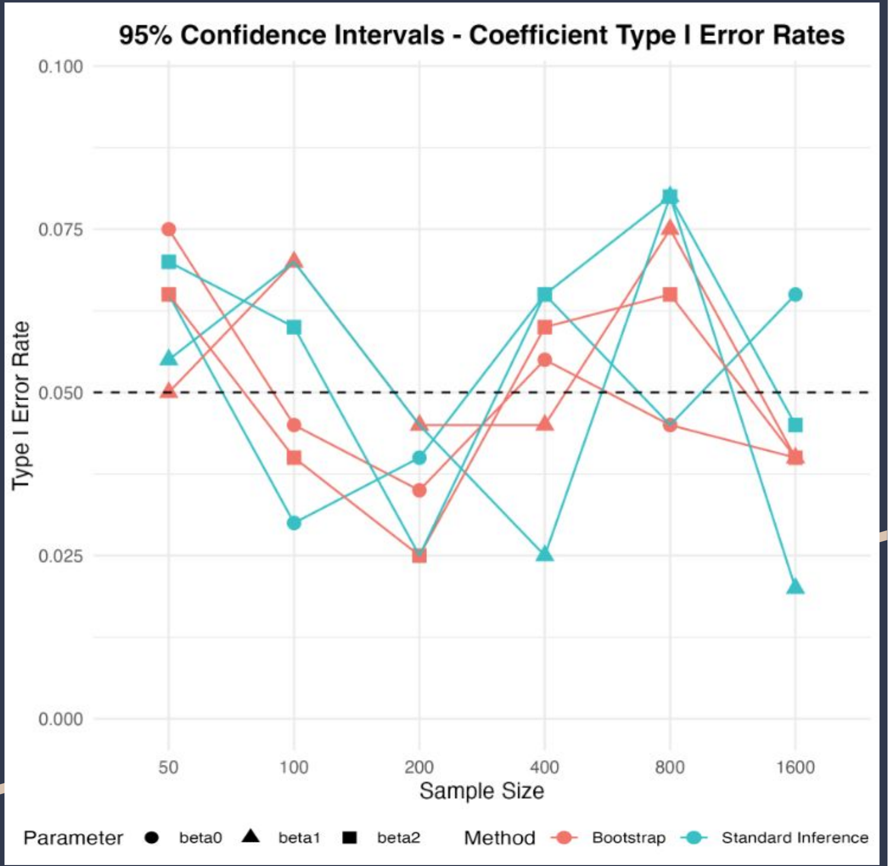
N	conf_level	expected_type1	std_type1_beta0	boot_type1_beta0	std_type1_beta1	boot_type1_beta1	std_type1_beta2	boot_type1_beta2	std_type1_transform	boot_type1_transform
50	0.8	0.2	0.15	0.175	0.17	0.18	0.18	0.15	0.025	0.175
50	0.9	0.1	0.09	0.095	0.07	0.105	0.135	0.1	0	0.12
50	0.95	0.05	0.065	0.075	0.055	0.05	0.07	0.065	0	0.08
100	0.8	0.2	0.175	0.17	0.205	0.195	0.185	0.14	0.005	0.17
100	0.9	0.1	0.095	0.085	0.065	0.095	0.09	0.07	0.005	0.105
100	0.95	0.05	0.03	0.045	0.07	0.07	0.06	0.04	0	0.07
200	0.8	0.2	0.155	0.14	0.17	0.16	0.19	0.185	0.02	0.205
200	0.9	0.1	0.105	0.1	0.08	0.105	0.105	0.075	0	0.095
200	0.95	0.05	0.04	0.035	0.045	0.045	0.025	0.025	0	0.055
400	0.8	0.2	0.175	0.185	0.175	0.17	0.175	0.16	0.005	0.155
400	0.9	0.1	0.065	0.07	0.095	0.11	0.085	0.06	0	0.08
400	0.95	0.05	0.065	0.055	0.025	0.045	0.065	0.06	0	0.065
800	0.8	0.2	0.225	0.2	0.17	0.165	0.185	0.165	0	0.16
800	0.9	0.1	0.11	0.09	0.09	0.095	0.12	0.06	0	0.1
800	0.95	0.05	0.045	0.045	0.08	0.075	0.08	0.065	0	0.065
1600	0.8	0.2	0.21	0.225	0.19	0.175	0.205	0.155	0.01	0.215
1600	0.9	0.1	0.06	0.055	0.125	0.15	0.105	0.085	0.005	0.12
1600	0.95	0.05	0.065	0.04	0.02	0.04	0.045	0.04	0	0.04



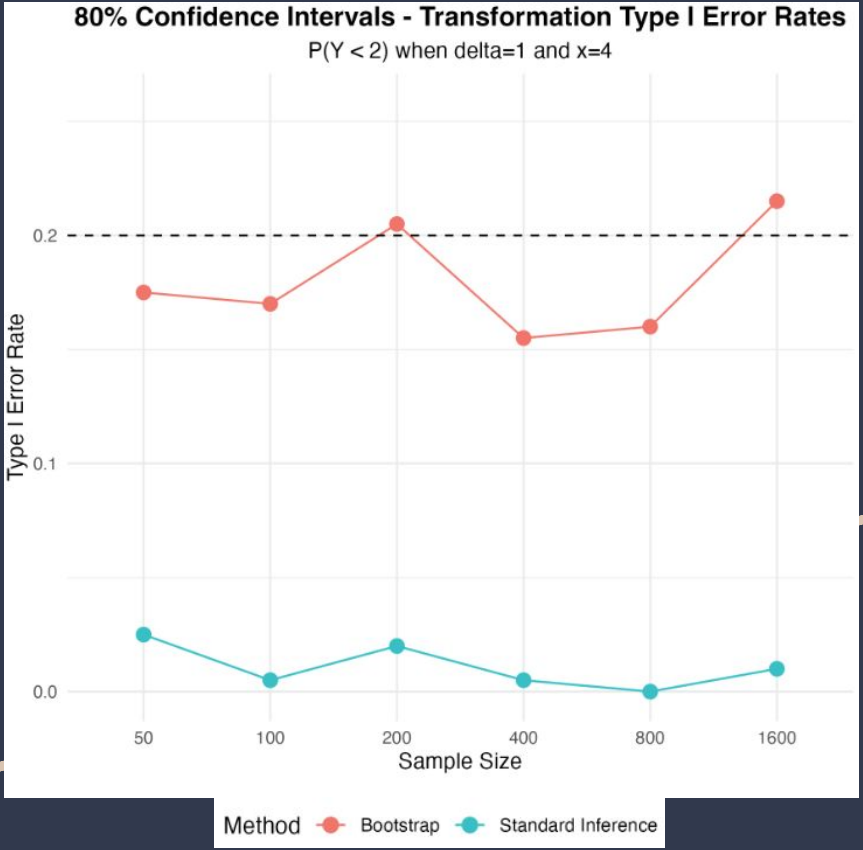
- Bootstrap
  - Beta 0
    - Linearly Increasing
    - Crosses 0.2 at 800
  - Beta 1
    - Peaks at 100
    - Not very variable
  - Beta 2
    - Peaks at 200
    - Results in lowest type 1 error
- Standard Inference
  - Higher type 1 error overall
  - Beta 0
    - Ends the highest
    - Peak at 800
  - Beta 1
    - Crosses at 100 - Drops after
  - Beta 2
    - Crosses at 1600
    - Similar to Beta 0 at max sample



- Bootstrap
  - Beta 0
    - Decreasing with increasing sample size
    - Never crosses significance threshold
  - Beta 1
    - More often above 0.1
    - Large increase between 800 & 1600
  - Beta 2
    - Consistently low
- Standard Inference
  - Error more centralized around 0.1 than Bootstrap
  - Beta 0
    - Variable
    - Highest error at 800
  - Beta 1
    - Consistent increase - peak at 1600
  - Beta 2
    - Close to 0.1 over most sizes

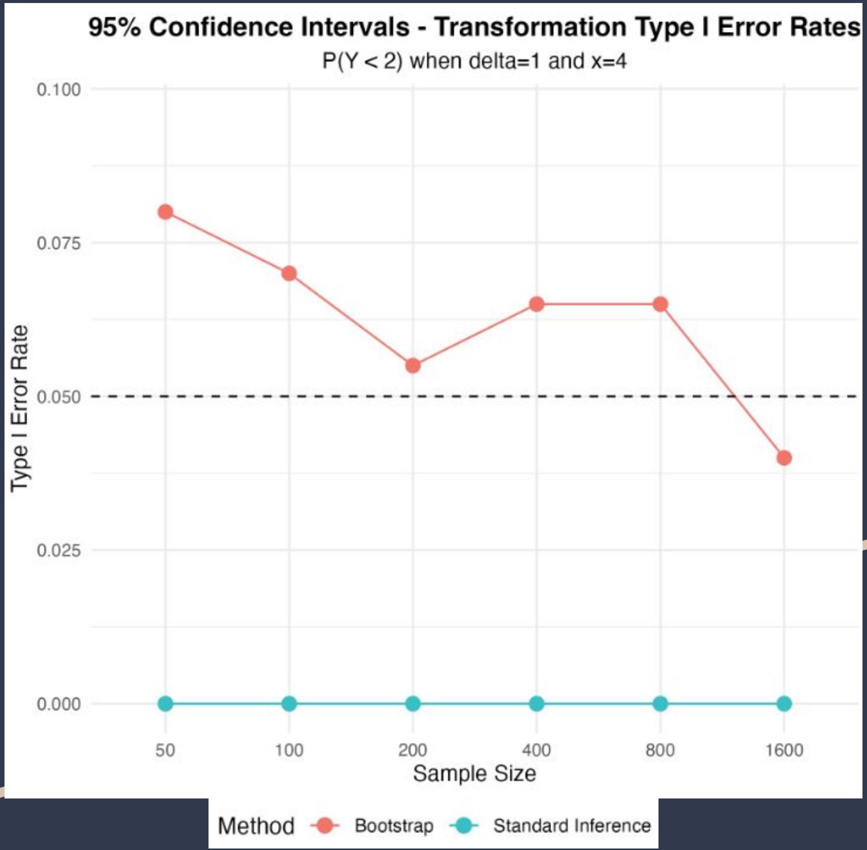


- All much more variable at 95%
- Bootstrap
  - All betas meet at about 0.04 at 1600 sample size
  - All points begin above 0.05 threshold
  - Beta 0
    - Least volatile
    - Type 1 error peak at n=50
  - Beta 1
    - Peak at n=800
  - Beta 2
    - Low at n=200
- Standard Inference
  - All points begin above 0.05 threshold
  - Beta 0
    - Mostly a steady increase
  - Beta 1
    - Joint highest error rate at n=800 ~ 0.075
    - Lowest at n=1600 ~ 0.02
  - Beta 2
    - Joint highest



- Standard Inference much lower than 0.2
- Bootstrap
  - Crosses 0.2 at n=200 and n=1600
  - More variable
- Standard Inference
  - Peaks at n=50

- Standard Inference again much lower
- Bootstrap
  - V-shape
  - Peaks at n=50 and n=1600
  - Low at 400
- Standard Inference
  - Peaks at n=100 and n=1600



- Standard inference again lower
- Bootstrap
  - Steadily decreasing
  - Starts highest at n=50
- Standard inference
  - Leads to almost horizontal type 1 error rates for the transformation