Simulation study

Simulation study: why we need it?

Target: Evaluate the performance of the methodology using the user-generated data sets from the true model.

1. We are not sure of the real performance of a newly proposed method.
2. With the generated data sets, we can test whether the method can reproduce the true parameters.

Simulation study: procedures

1. specify your model
2. give true parameter values
3. generate data sets (usually 100 replications)
4. implement your method
5. summarize the estimation results (Bias, RMS, etc.)

Simulation study: example

1. Specify my model

Let y_i denotes the count outcome that follows $Pois(\mu_i)$, and the loglinear model can be expressed as follows:

$$y_i \sim Pois(\mu_i)$$
$$\log(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$$

where x_1 and x_2 denote observed covariates, and β_0, β_1, β_2 are unknown parameters.

Simulation study: example

2. Give true parameter values

$$y_i \sim Pois(\mu_i)$$
$$\log(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$$

- sample size: $n = 300$
- number of covariates: $p = 2$
- parameter true values: $\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = -1$
- number of replications: $n_{rep} = 10$

Simulation study: example

3. Generate data sets

$$y_i \sim Pois(\mu_i)$$
$$\log(\mu_i) = 0.5 + x_{1i} - x_{2i}$$

1. generate x_{1i}, x_{2i} from $N(0, 1)$ (try other distributions!)
2. calculate μ_i based on the true model
3. generate y_i from poisson distribution with parameter μ_i
4. replicate for $n_{rep} = 10$ times and save

see examples/simu/glm_count.R generate data part

Simulation study: example

4. Implement your method

see examples/simu/glm_count.R estimation part

5. Summarize the estimation results

see examples/simu/glm_count.R summary part