Statistical Power for Logistic Regression

Haiyan Liu

UNIVERSITY OF NOTRE DAME

QSG 2015S

Feb 26, 2014

Outline

- Introduction
  - Power
  - Logistic regression
- Hypothesis testing
  - Wald test
  - Score test
  - Likelihood ratio test
- Power calculation
- WebPower and future directions

Why power?

- Power: the probability of rejecting the null hypothesis given that the null hypothesis is false.

$$\text{power} = P(\text{reject null} | \text{null is false})$$

- two forms of power analysis are most useful:
  - the determination of the sample size n required to attain a specified degree of power
  - the determination of the power to detect the hypothesized effect size.

Factors that affect power

The power of a hypothesis test is affected by at least three factors:

1. Significance level (α): the degree to which H_0 is false. The higher the significance level, the higher the power of the test, when other factors are fixed..

Factors that affect power

The power of a hypothesis test is affected by at least three factors:

1. Significance level (α): the degree to which H_0 is false. The higher the significance level, the higher the power of the test, when other factors are fixed..
2. Sample size (n): Other things being equal, the greater the sample size, the greater the power of the test.

Factors that affect power

The power of a hypothesis test is affected by at least three factors:

1. Significance level (α): the degree to which H_0 is false. The higher the significance level, the higher the power of the test, when other factors are fixed..
2. Sample size (n): Other things being equal, the greater the sample size, the greater the power of the test.
3. Effect size (ES): the discrepancy between the the "true" value of the parameter being tested and the value specified in the null hypothesis. The greater the effect size, the greater power of the test.

Factors that affect power

The power of a hypothesis test is affected by at least three factors:

1. Significance level (α): the degree to which H_0 is false. The higher the significance level, the higher the power of the test, when other factors are fixed..
2. Sample size (n): Other things being equal, the greater the sample size, the greater the power of the test.
3. Effect size (ES): the discrepancy between the the "true" value of the parameter being tested and the value specified in the null hypothesis. The greater the effect size, the greater power of the test.
4. Others: the tests methods, distribution of predictors, missing data ....

Logistic Regression

Description of the model

- $Y|X = x \sim \text{Bernoulli}(\pi(x))$

Logistic Regression

Description of the model

- $Y|X = x \sim \text{Bernoulli}(\pi(x))$
- $\eta(x) = b_0 + x_1 b_1 + \cdots x_p b_p$

Logistic Regression

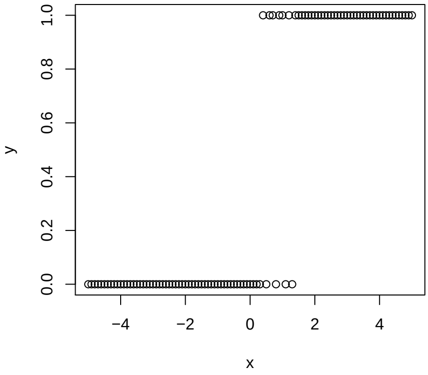
Description of the model

- $Y|X = x \sim \text{Bernoulli}(\pi(x))$
- $\eta(x) = b_0 + x_1 b_1 + \cdots x_p b_p$
- $\pi(x) = \frac{e^{\eta(x)}}{1+e^{\eta(x)}} \iff \log \frac{\pi(x)}{1-\pi(x)} = b_0 + x_1 b_1 + \cdots x_p b_p$

Logistic Regression

Description of the model

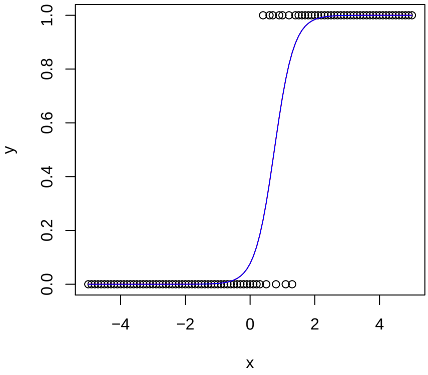
- $Y|X = x \sim \text{Bernoulli}(\pi(x))$
- $\eta(x) = b_0 + x_1 b_1 + \cdots x_p b_p$
- $\pi(x) = \frac{e^{\eta(x)}}{1+e^{\eta(x)}} \iff \log \frac{\pi(x)}{1-\pi(x)} = b_0 + x_1 b_1 + \cdots x_p b_p$



Logistic Regression

Description of the model

- $Y|X = x \sim \text{Bernoulli}(\pi(x))$
- $\eta(x) = b_0 + x_1 b_1 + \cdots x_p b_p$
- $\pi(x) = \frac{e^{\eta(x)}}{1+e^{\eta(x)}} \iff \log \frac{\pi(x)}{1-\pi(x)} = b_0 + x_1 b_1 + \cdots x_p b_p$



Estimation

- Likelihood function/log-likelihood

$$
\begin{aligned}
L(b_0, b_1, \cdots, b_p; y, x) &= \Pi_{i=1}^{N} \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \\
l(b_0, b_1, \cdots, b_p; y, x) &= \sum_{i=1}^{N} \{ y_i \pi(x_i) + (1 - y_i)[1 - \pi(x_i)] \}
\end{aligned}
$$

Estimation

- Likelihood function/log-likelihood

$$
\begin{aligned}
L(b_0, b_1, \cdots, b_p; y, x) &= \Pi_{i=1}^{N} \pi(x_i)^{y_i}[1 - \pi(x_i)]^{1-y_i} \\
l(b_0, b_1, \cdots, b_p; y, x) &= \sum_{i=1}^{N} \{y_i \pi(x_i) + (1 - y_i)[1 - \pi(x_i)]\}
\end{aligned}
$$

- Newton-Raphson method

Estimation

- Likelihood function/log-likelihood

$$L(b_0, b_1, \cdots, b_p; y, x) = \Pi_{i=1}^{N} \pi(x_i)^{y_i}[1 - \pi(x_i)]^{1-y_i}$$

$$l(b_0, b_1, \cdots, b_p; y, x) = \sum_{i=1}^{N} \{y_i \pi(x_i) + (1 - y_i)[1 - \pi(x_i)]\}$$

- Newton-Raphson method: suppose $\hat{\beta}$ is MLE of β, when $n \rightarrow \infty$
  1. $\hat{\beta} \sim N(\beta', \Sigma)$(asymptotically normal)
  2. $\beta' = \beta$(consistency)
  3. $\Sigma = I^{-1}$ where I is the Fisher-information matrix. $I_{jk} = -E(\frac{\partial^2 l(\beta)}{\partial \beta_j \beta_k})$. Particularly, in one dimension, $\sigma^2 = cov(\hat{\beta}) = I^{-1} = E[-\frac{\partial^2 l(\beta)}{\partial \beta^2}]$.

Hypothesis testing

In simple logistic regression, let $\beta = b_1$(slope)

$$H_0 : \beta = 0 \qquad \text{vs} \qquad H_1 : \beta \neq 0$$

Hypothesis testing

In simple logistic regression, let $\beta = b_1$(slope)

$$H_0 : \beta = 0 \qquad \text{vs} \qquad H_1 : \beta \neq 0$$

- Effect size index:

$$b_1 = \log \frac{p_1/(1-p_1)}{p_0/(1-p_0)}$$

where $p_0 = Pr(Y = 1|H_0)$ and $p_1 = Pr(Y = 1|H_1 \& X = 1)$.

- Contional on H_1 is true, but we fit the H_0, we get

$$b_0^* = \log \frac{\pi}{1-\pi}$$

where $\pi = E(\pi(x))|_{H_1}) = E_{X\sim F}(\frac{1}{1+\exp(-b_0-b_1X)})$.

Three test statistics for $H_0 : \beta = 0$

- Wald test: $Z_w = \dfrac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\hat{\beta}}}} \rightsquigarrow N(0,1)$
- Score test : $Z_s = \dfrac{\frac{\partial l(\beta)}{\partial \beta}|_{\beta=0}}{\sqrt{-E(\frac{\partial^2 l(\beta)}{\partial \beta^2})|_{\beta=0}}} \rightsquigarrow N(0,1)$
- Likelihood ratio test: $S_{LR} = 2(l_1 - l_0) \rightsquigarrow \mathcal{X}^2_{df=1}$

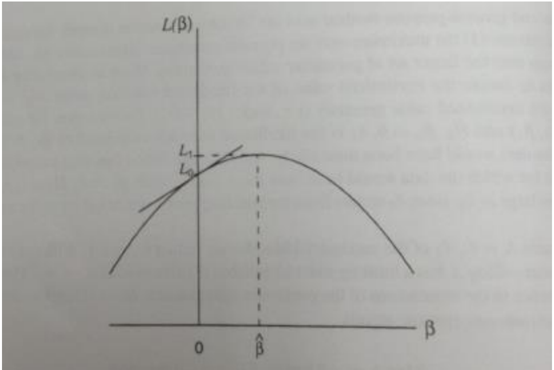


Figure 1.1 Log-likelihood function and information used in three tests of H_0: $\beta = 0$.

Power Calculation

Hypothesis: $H_0 : \beta = 0$ vs $H_1 : \beta \neq 0$:

$$\text{Power} = P(H_0 \text{ is rejected} | H_1 \text{ is true})$$

- Under H_0 :

$$Z_w = \frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\hat{\beta}}}} \rightsquigarrow N(0,1)$$

Power Calculation

Hypothesis: $H_0 : \beta = 0$ vs $H_1 : \beta \neq 0$:

$$\text{Power} = P(H_0 \text{ is rejected} | H_1 \text{ is true})$$

- Under H_0 :

$$Z_w = \frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\hat{\beta}}}} \rightsquigarrow N(0,1)$$

- Under H_1 :

$$\Gamma = \frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}} \rightsquigarrow N(\beta,1)$$

Power Calculation

Hypothesis: $H_0 : \beta = 0$ vs $H_1 : \beta \neq 0$:

$$\text{Power} = P(H_0 \text{ is rejected} | H_1 \text{ is true})$$

- Under H_0 :

$$Z_w = \frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\hat{\beta}}}} \rightsquigarrow N(0,1)$$

- Under H_1 :

$$\Gamma = \frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}} \rightsquigarrow N(\beta,1)$$

- Bickel, Doksum(2001), Demidenko(2004), by Slusky theorem

$$var(\hat{\beta})|_{\hat{\beta}} \to var(\hat{\beta})|_{\beta}$$

Power(Demidenko, 2007)

- Power:= $P(H_0 \text{ is rejected} | H_1 \text{ is true})$

Power(Demidenko, 2007)

- Power:= $P(H_0 \text{ is rejected} | H_1 \text{ is true})$
- power= $P\left(\left|\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\hat{\beta}}}}\right| > Z_{1-\frac{\alpha}{2}} | H_1 \text{ is true}\right)$

Power(Demidenko, 2007)

- Power:= $P(H_0 \text{ is rejected} | H_1 \text{ is true})$
- power= $P(|\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\hat{\beta}}}}| > Z_{1-\frac{\alpha}{2}} | H_1 \text{ is true})$
- power≈ $P(|\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}}| > Z_{1-\frac{\alpha}{2}} | H_1 \text{ is true})$

Power(Demidenko, 2007)

- Power:= $P(H_0 \text{ is rejected}|H_1 \text{ is true})$
- power= $P(|\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\hat{\beta}}}}| > Z_{1-\frac{\alpha}{2}}|H_1 \text{ is true})$
- power$\approx P(|\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}}| > Z_{1-\frac{\alpha}{2}}|H_1 \text{ is true})$
- power$\approx P(\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}} > Z_{1-\frac{\alpha}{2}}|H_1 \text{ is true}) + P(\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}} < -Z_{1-\frac{\alpha}{2}}|H_A \text{ is true})$

Power(Demidenko, 2007)

- Power:= $P(H_0 \text{ is rejected}|H_1 \text{ is true})$
- power= $P(|\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\hat{\beta}}}}| > Z_{1-\frac{\alpha}{2}}|H_1 \text{ is true})$
- power≈ $P(|\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}}| > Z_{1-\frac{\alpha}{2}}|H_1 \text{ is true})$
- power≈ $P(\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}} > Z_{1-\frac{\alpha}{2}}|H_1 \text{ is true}) + P(\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}} < -Z_{1-\frac{\alpha}{2}}|H_A \text{ is true})$
- Power≈ $P(\frac{\hat{\beta}-\beta}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}} > Z_{1-\frac{\alpha}{2}} - \frac{\beta}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}}|H_1 \text{ is true}) + P(\frac{\hat{\beta}-\beta}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}} < -Z_{1-\frac{\alpha}{2}} - \frac{\beta}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}}|H_1 \text{ is true})$

Power(Demidenko, 2007)

- Power:= $P(H_0 \text{ is rejected} | H_1 \text{ is true})$
- power= $P(|\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\hat{\beta}}}}| > Z_{1-\frac{\alpha}{2}} | H_1 \text{ is true})$
- power≈ $P(|\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}}| > Z_{1-\frac{\alpha}{2}} | H_1 \text{ is true})$
- power≈ $P(\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}} > Z_{1-\frac{\alpha}{2}} | H_1 \text{ is true}) + P(\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}} < -Z_{1-\frac{\alpha}{2}} | H_A \text{ is true})$
- Power≈ $P(\frac{\hat{\beta}-\beta}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}} > Z_{1-\frac{\alpha}{2}} - \frac{\beta}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}} | H_1 \text{ is true}) + P(\frac{\hat{\beta}-\beta}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}} < -Z_{1-\frac{\alpha}{2}} - \frac{\beta}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}} | H_1 \text{ is true})$
- Power≈ $\Phi(-Z_{1-\frac{\alpha}{2}} + \frac{\beta}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}}) + \Phi(-Z_{1-\frac{\alpha}{2}} - \frac{\beta}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}})$

Power(Demidenko, 2007)

- Power:= $P(H_0$ is rejected$|H_1$ is true$)$
- power= $P(|\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\hat{\beta}}}}| > Z_{1-\frac{\alpha}{2}}|H_1$ is true$)$
- power≈ $P(|\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}}| > Z_{1-\frac{\alpha}{2}}|H_1$ is true$)$
- power≈ $P(\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}} > Z_{1-\frac{\alpha}{2}}|H_1$ is true$) + P(\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}} < -Z_{1-\frac{\alpha}{2}}|H_A$ is true$)$
- Power≈ $P(\frac{\hat{\beta}-\beta}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}} > Z_{1-\frac{\alpha}{2}} - \frac{\beta}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}}|H_1$ is true$) +$ $P(\frac{\hat{\beta}-\beta}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}} < -Z_{1-\frac{\alpha}{2}} - \frac{\beta}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}}|H_1$ is true$)$
- Power≈ $\Phi(-Z_{1-\frac{\alpha}{2}} + \frac{\beta}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}}) + \Phi(-Z_{1-\frac{\alpha}{2}} - \frac{\beta}{\sqrt{var(\hat{\beta})|_{\beta=\beta}}})$

What's $var(\hat{\beta})|_{\beta=\beta}$?

- Log-likelihood

$$l(b_0, b_1; y, x) = \sum_{i=1}^{N} y_i \log \pi(x_i) + (1 - y_i) \log(1 - \pi(x_i))$$

What's $var(\hat{\beta})|_{\beta=\beta}$?

- Log-likelihood

$$l(b_0, b_1; y, x) = \sum_{i=1}^{N} y_i \log \pi(x_i) + (1 - y_i) \log(1 - \pi(x_i))$$

- Fisher-information:

$$\frac{\partial l_i(b_0, b_1)}{\partial b_0^2} = -\pi(x_i)[1 - \pi(x_i)]$$

$$\frac{\partial l_i^2(b_0, b_1)}{\partial b_1^2} = -x_i^2 \pi(x_i)(1 - \pi(x_i))$$

$$\frac{\partial l_i^2(b_0, b_1)}{\partial b_0 b_1} = -x_i \pi(x_i)[1 - \pi(x_i))]$$

$$I = n \begin{bmatrix} E\pi(x_i)[1 - \pi(x_i)] & Ex_i\pi(x_i)(1 - \pi(x_i)) \\ E\pi(x_i)[1 - \pi(x_i)] & Ex_i^2\pi(x_i)(1 - \pi(x_i)) \end{bmatrix} = \begin{bmatrix} I_{00} \\ I_{01} \end{bmatrix}$$

$$I^{-1} = \frac{1}{n(I_{00} I_{11} - I_{01}^2)} \begin{bmatrix} I_{11} & -I_{01} \\ -I_{01} & I_{00} \end{bmatrix}$$

What's $var(\hat{\beta})|_{\beta=\beta}$?

- Log-likelihood

$$l(b_0, b_1; y, x) = \sum_{i=1}^{N} y_i \log \pi(x_i) + (1 - y_i) \log(1 - \pi(x_i))$$

- Fisher-information:

$$\begin{aligned}
\frac{\partial l_i(b_0, b_1)}{\partial b_0^2} &= -\pi(x_i)[1 - \pi(x_i)] \\
\frac{\partial l_i^2(b_0, b_1)}{\partial b_1^2} &= -x_i^2 \pi(x_i)(1 - \pi(x_i)) \\
\frac{\partial l_i^2(b_0, b_1)}{\partial b_0 b_1} &= -x_i \pi(x_i)[1 - \pi(x_i))] \\
I &= n \begin{bmatrix} E\pi(x_i)[1 - \pi(x_i)] & Ex_i\pi(x_i)(1 - \pi(x_i)) \\ E\pi(x_i)[1 - \pi(x_i)] & Ex_i^2\pi(x_i)(1 - \pi(x_i)) \end{bmatrix} = \begin{bmatrix} I_{00} \\ I_{01} \end{bmatrix} \\
I^{-1} &= \frac{1}{n(I_{00} I_{11} - I_{01}^2)} \begin{bmatrix} I_{11} & -I_{01} \\ -I_{01} & I_{00} \end{bmatrix}
\end{aligned}$$

- $var(\hat{\beta}) = \frac{I_{00}}{n(I_{00} I_{11} - I_{01}^2)} = v_1/n$ in which b_0 and b_1 are given by H_1

Variance with correction

- $var(\hat{\beta}) = \frac{I_{00}}{n(I_{00}I_{11}-I_{01}^2)} = v_1/n$ in which b_0 and b_1 are given by H_1.
- Under H_1 :

$$\Gamma = \frac{\hat{\beta}-\beta}{\sqrt{\frac{v_1}{n}}} \rightsquigarrow N(0,1)$$

- With variance correction,

$$\Gamma = \frac{\sqrt{n}(\hat{\beta}-\beta)}{\sqrt{v_1}} \rightsquigarrow N(0, \sqrt{\frac{av_0+(1-a)v_1}{v_1}})$$

where $v_0$is the variance of $\hat{\beta}_1$for H_0 with $b_0 = \lg \frac{\mu}{1-\mu}$, with

$$\mu = E_X(\frac{\exp(b_0+b_1X)}{1+\exp(b_0+b_1X)})$$

- Thus under the $H_1 : \Gamma = \frac{\hat{\beta}-\beta}{\sqrt{\frac{av_0+(1-a)v_1}{n}}} \sim N(0,1)$
- Power$= \Phi(-Z_{1-\frac{\alpha}{2}} + \sqrt{n}\beta/\sqrt{av_0+(1-a)v_1}) + \Phi(-Z_{1-\frac{\alpha}{2}} - \sqrt{n}\beta/\sqrt{av_0+(1-a)v_1})$.

Variances

In the power expression:

1. $v_0 = \frac{I_{00}}{I_{00} I_{11} - I_{01}^2}$ evaluated under $H_0 : b_{0*} = \log$
   1.1 and $v_1 = \frac{I_{00}}{I_{00} I_{11} - I_{01}^2}$ evaluated under H_1
2. Under H_1 :
   2.1 $I_{00} = E\pi(x)[1 - \pi(x)]\backslash$
   2.2 $I_{11} = Ex^2\pi(x)[1 - \pi(x)]$
   2.3 $I_{01} = Ex\pi(x)[1 - \pi(x)]$
3. $X \sim Bernoulli(B)$

$$I_{00} =$$

Distribution of the covariate

- Bernoulli(p)
- Exponential(λ)
- Lognormal(μ, σ)
- Normal(μ, σ)
- Poisson(λ)
- Uniform(L, R)

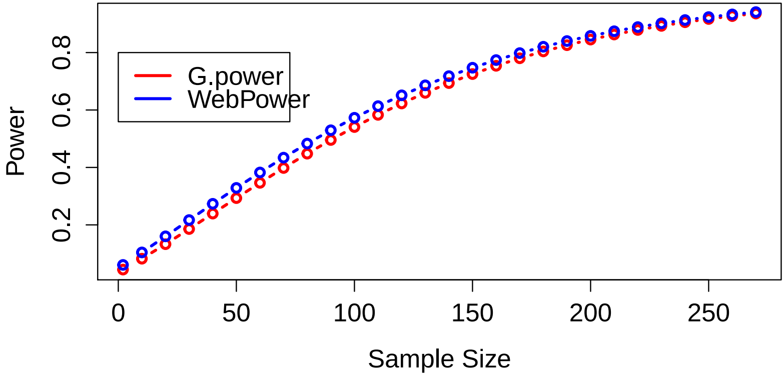
WebPower

- Input model:
  - $p_0 = Pr(Y = 1|H_0)$ and $p_1 = Pr(Y = 1|H_A, X = 1)$
  - significant level α
  - either n or power
  - family: distribution of the predictor

`http://webpower.psychstat.org/`

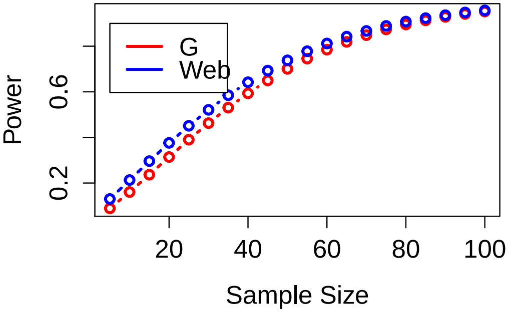
WebPower VS G-power

$p_0 = 0.2, p_1 = 0.3, \alpha = 0.05,\ X \sim N(0,1)$



WebPower VS G-power

$p_0 = 0.2, p_1 = 0.3, \alpha = 0.05,\ X \sim U[0, 5]$



Future directions

- Compare the three test statistics .

Future directions

- Compare the three test statistics .
- Extend the procedures to multiple logistic regression.

Future directions

- Compare the three test statistics .
- Extend the procedures to multiple logistic regression.
- Validate out results through MCMC.

Future directions

- Compare the three test statistics .
- Extend the procedures to multiple logistic regression.
- Validate out results through MCMC.

Acknowledgment

- Johnny Zhang
- Ke-Hai Yuan
- My colleagues
  - Yujiao, Megan, Han, Ge, Miao, Asangus