

**Class 6:
Chapter 8**

Min Lu

Object:

Wald Confidence
interval

Profile Likelihood
Confidence Interval

R Example

Class 6: Chapter 8

EPH 705

Min Lu

Division of Biostatistics
University of Miami

Spring 2017

**Class 6:
Chapter 8**

Min Lu

Object:

Wald Confidence
interval
Profile Likelihood
Confidence Interval

R Example

- 1 Object:
 - Wald Confidence interval
 - Profile Likelihood Confidence Interval

- 2 R Example

The lower endpoint of the 95% confidence interval is:

$$\text{Lower endpoint} = \hat{\beta} - z_{1-\alpha/2} \times SE_{\hat{\beta}},$$

and the upper endpoint of the 95% confidence interval is:

$$\text{Upper endpoint} = \hat{\beta} + z_{1-\alpha/2} \times SE_{\hat{\beta}}.$$

Profile Likelihood Confidence Interval

Class 6: Chapter 8

Min Lu

Object:

Wald Confidence
interval

Profile Likelihood
Confidence Interval

R Example

Wald-type CI may work poorly if the distribution of the parameter estimator is markedly skewed or if the standard error is a poor estimate of the standard deviation of the estimator. Since the standard errors in GLM's are based on asymptotic variances obtained from the information matrix, Wald CI's may perform poorly for small to moderate sample sizes. Profile likelihood confidence intervals don't assume normality of the estimator and appear to perform better for small samples sizes than Wald CI'S. They are, nonetheless, still based on an asymptotic approximation – the asymptotic chi-square distribution of the log likelihood ratio test statistic.

Profile Likelihood Confidence Interval

Class 6: Chapter 8

Min Lu

Object:

Wald Confidence
interval

Profile Likelihood
Confidence Interval

R Example

Assume that each i is associated either with a single Bernoulli trial or with n_i independent identically distributed trials, where the observation Y_i is the number of successes observed (the sum of the individual Bernoulli-distributed random variables), and hence follows a binomial distribution: $Y_i \sim \text{Bin}(n_i, p_i)$, for $i = 1, \dots, n$

$$\Pr(Y_i = y) = \binom{n_i}{y} p_i^y (1 - p_i)^{n_i - y}$$

Logistic regression for binomial outcome

$$\Pr(Y_i = y \mid \mathbf{X}_i) = \binom{n_i}{y} \left(\frac{1}{1 + e^{-\beta \cdot \mathbf{X}_i}} \right)^y \left(1 - \frac{1}{1 + e^{-\beta \cdot \mathbf{X}_i}} \right)^{n_i - y}.$$

Profile Likelihood Confidence Interval

Class 6: Chapter 8

Min Lu

Object:

Wald Confidence
interval

Profile Likelihood
Confidence Interval

R Example

The solution to a Logistic Regression problem is the set of parameters β that maximizes the likelihood of the data, which is expressed as the product of the predicted probabilities of the N individual observations.

$$f(y_1, y_2, \dots, y_n | \beta) = f(y_1 | \beta) \times f(y_2 | \beta) \times \dots \times f(y_n | \beta)$$

this same function will be called the likelihood:

$$\mathcal{L}(\beta; y_1, \dots, y_n) = f(y_1, y_2, \dots, y_n | \beta) = \prod_{i=1}^n f(y_i | \beta)$$

In practice the algebra is often more convenient when working with the natural logarithm of the likelihood function, called the log-likelihood:

$$\ln \mathcal{L}(\theta; y_1, \dots, y_n) = \sum_{i=1}^n \ln f(y_i | \beta),$$

Class 6:
Chapter 8

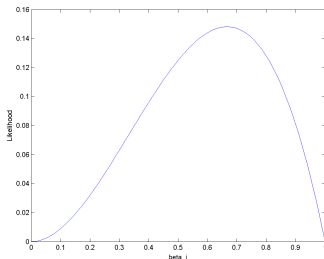
Min Lu

Object:

Wald Confidence
intervalProfile Likelihood
Confidence Interval

R Example

$$\begin{aligned}\mathcal{L}(\beta; y_1, \dots, y_n) &= \prod_{i=1}^n \binom{n_i}{y_i} p_i^{y_i} (1 - p_i)^{n_i - y_i} \\ &= \prod_{i=1}^n \binom{n_i}{y_i} \left(\frac{1}{1 + e^{-\beta \cdot \mathbf{X}_i}} \right)^{y_i} \left(1 - \frac{1}{1 + e^{-\beta \cdot \mathbf{X}_i}} \right)^{n_i - y_i}.\end{aligned}$$



Marginal Profile Likelihood Confidence Interval

Class 6: Chapter 8

Min Lu

Object:

Wald Confidence
interval

Profile Likelihood
Confidence Interval

R Example

Suppose in a two-parameter problem we wish to carry out a hypothesis test for β_1 .

$$H_0 : \beta_1 = \beta_{01}$$

$$H_0 : \beta_1 \neq \beta_{01}$$

To carry out this test we need to compare two log-likelihoods,

$$LR_{\beta_1|\beta_2} = 2\log \frac{L(\hat{\beta}_1, \hat{\beta}_2)}{L(\hat{\beta}_{01}, \hat{\beta}_2)} \sim \chi_1^2$$

Here $\hat{\beta}_1$ and $\hat{\beta}_2$ are the MLEs (maximum likelihood estimates). We reject the null hypothesis at $\alpha = .05$ if

$$LR_{\beta_1|\beta_2} > \chi_1^2(0.95)$$

If we flip the inequality around we get the “acceptance” region, all values of the likelihood ratio statistic that would cause us to not reject the null hypothesis. The acceptance region can be written as follows.

$$\beta_1 : \log(L(\beta_1, \hat{\beta}_2)) \geq \log(L(\hat{\beta}_1, \hat{\beta}_2)) - \frac{1}{2}\chi_1^2(0.95)$$

Class 6:
Chapter 8

Min Lu

Object:

Wald Confidence
intervalProfile Likelihood
Confidence Interval

R Example

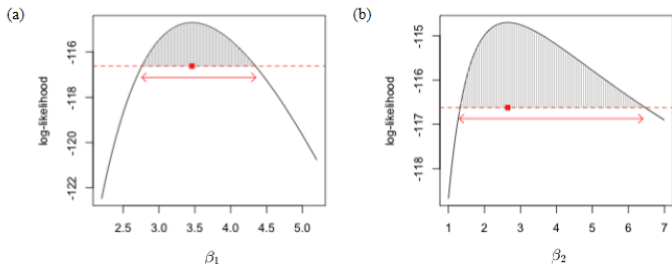


Fig. 2 Plot of the profile log-likelihood for (a) β_1 and (b) β_2 showing the maximum likelihood estimate (red square ■) and the 95% profile likelihood confidence interval (red arrow \leftrightarrow). Shaded regions are the acceptance regions obtained from the likelihood ratio test.

Western Collaborative Group Study data

Class 6: Chapter 8

Min Lu

Object:

Wald Confidence
interval
Profile Likelihood
Confidence Interval

R Example

Data: The Western Collaborative Group Study (WCGS), a prospective cohort study, recruited middle-aged men (ages 39 to 59) who were employees of 10 California companies and collected data on 3154 individuals during the years 1960-1961. These subjects were primarily selected to study the relationship between behavior pattern and the risk of coronary heart disease (CHD). A number of other risk factors were also measured.

variable name	discreption
id	Subject ID:
age0	Age: age in years
height0	Height: height in inches
weight0	Weight: weight in pounds
sbp0	Systolic blood pressure: mm Hg
dbp0	Diastolic blood pressure: mm Hg
chol0	Cholesterol: mg/100 ml
behpat0	Behavior pattern:
ncigs0	Smoking: Cigarettes/day
dibpat0	Dichotomous behavior pattern: 0 = Type B; 1 = Type A
chd69	Coronary heart disease event: 0 = none; 1 = yes
typechd	to be done
time169	Observation (follow up) time: Days
arcus0	Corneal arcus: 0 = none; 1 = yes

Object:

Wald Confidence
interval
Profile Likelihood
Confidence Interval

R Example

Logistic Regression: Overdispersion

```
wcgs <- read.csv("wcgs.csv")[, -1]
object <- glm(arcus0 ~ age0 + height0 + dibpat0 + typechd + chd69, data = wcgs, family <- binomial(link = "logit"))
object.ov <- glm(arcus0 ~ age0 + height0 + dibpat0 + typechd + chd69, data = wcgs, family <- quasibinomial(link = "logit"))
summary(object.ov)

##
## Call:
## glm(formula = arcus0 ~ age0 + height0 + dibpat0 + typechd + chd69,
##      family = family <- quasibinomial(link = "logit"), data = wcgs)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4286  -0.8460  -0.7136   1.2870   1.8910
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.441165   1.198937  -5.372 8.34e-08 ***
## age0         0.072270   0.007209  10.025 < 2e-16 ***
## height0     0.030573   0.015924   1.920 0.05495 .
## dibpat0     0.099894   0.080396   1.243 0.21414
## typechd    -0.310570   0.169692  -1.830 0.06732 .
## chd69       0.821205   0.310457   2.645 0.00821 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.00171)
##
##      Null deviance: 3843.1 on 3151 degrees of freedom
## Residual deviance: 3720.0 on 3146 degrees of freedom
##      (2 observations deleted due to missingness)
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

Object:

Wald Confidence
interval
Profile Likelihood
Confidence Interval

R Example

Logistic Regression: confidence interval

```
# Wald 95% confidence interval for coefficients
bhat <- summary(object)$coefficients[, 1]
bhatCI <- cbind(bhat, bhat - 1.96 * summary(object)$coefficients[, 2], bhat + 1.96 * summary(object)$coefficients[,
2])
colnames(bhatCI) <- c("coefficients", "LL", "UL")
head(bhatCI)

##              coefficients              LL              UL
## (Intercept) -6.44116465 -8.7890750415 -4.09325427
## age0         0.07226995  0.0581520387  0.08638786
## height0      0.03057285 -0.0006111515  0.06175686
## dibpat0      0.09989390 -0.0575477045  0.25733550
## typechd     -0.31056952 -0.6428822972  0.02174325
## chd69        0.82120487  0.2132290864  1.42918065

# profile-likelihood confidence intervals (95%) for coefficients
confint(object, level = 0.95)

## Waiting for profiling to be done...

##              2.5 %              97.5 %
## (Intercept) -8.8001790147 -4.10299558
## age0         0.0581917407  0.08643478
## height0      -0.0005501709  0.06183668
## dibpat0      -0.0575529014  0.25738853
## typechd     -0.6481796134  0.01819155
## chd69        0.2121635558  1.43017949
```

Object:

Wald Confidence
interval
Profile Likelihood
Confidence Interval

R Example

Logistic Regression: confidence interval

```
# 95% confidence interval for predicted values
yhat <- predict(object, se.fit = T, type = "response")
yhatCI <- cbind(yhat$fit, yhat$fit - 1.96 * yhat$se.fit, yhat$fit + 1.96 * yhat$se.fit)
colnames(yhatCI) <- c("pi", "LL", "UL")
head(yhatCI)
```

```
##           pi           LL           UL
## 1 0.3616465 0.3263802 0.3969128
## 2 0.2376090 0.2130966 0.2621214
## 3 0.2147884 0.1923102 0.2372666
## 4 0.1979522 0.1739798 0.2219245
## 5 0.6162112 0.5252631 0.7071593
## 6 0.2573015 0.2312861 0.2833170
```

Object:

Wald Confidence
interval
Profile Likelihood
Confidence Interval

R Example

Logistic Regression: Likelihood ratio and Wald test

```
##Likelihood ratio test of all betas equal to zero except intercept
LLR <- (object$null.deviance - object$deviance)
DF <- object$df.null - object$df.residual
## p_value <0.05 means coefficients are significantly diff from 0
pchisq(LLR, df = DF, lower.tail = FALSE)
```

```
## [1] 7.024307e-25
# Wald test to test coefficients of age0 (2) and height0 (3)
choose <- c(2, 3)
# variance-covariance matrix of the coefficients
V <- vcov(object)
Wald <- bhat[choose] %*% solve(V[choose, choose]) %*% bhat[choose]
## p_value <0.05 means coefficients are significant
pchisq(sum(Wald), df = length(choose), lower.tail = FALSE)
```

```
## [1] 1.059148e-22
```

Object:

Wald Confidence
intervalProfile Likelihood
Confidence Interval

R Example

Logistic Regression: Nested Model

```
object1 <- glm(arcus0 ~ age0 + height0 + dibpat0, data = wcgs, family = binomial(link = "logit"))
object2 <- glm(arcus0 ~ age0 + height0 + dibpat0 + typechd + chd69, data = wcgs, family = binomial(link = "logit"))
anv <- anova(object1, object2)
anv

## Analysis of Deviance Table
##
## Model 1: arcus0 ~ age0 + height0 + dibpat0
## Model 2: arcus0 ~ age0 + height0 + dibpat0 + typechd + chd69
##   Resid. Df Resid. Dev Df Deviance
## 1       3148      3728.2
## 2       3146      3720.0  2    8.2357
## p_value >0.05 Model 2 significantly improves model fit
pchisq(anv$Deviance[2], df = anv$Df[2], lower.tail = FALSE)

## [1] 0.0162795
```

Checkout distribution of the outcome

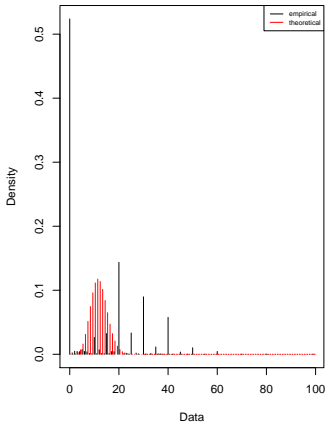
```
library('fitdistrplus')  
plot(fitdist(wcgs$ncigs0, "pois"))
```

Object:

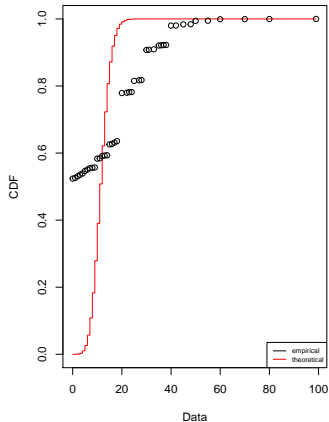
Wald Confidence
interval
Profile Likelihood
Confidence Interval

R Example

Emp. and theo. distr.



Emp. and theo. CDFs



**Class 6:
Chapter 8**

Min Lu

Object:

Wald Confidence
interval

Profile Likelihood
Confidence Interval

R Example

Use the SAS example and conduct analysis using R.

Take home exercise

Class 6: Chapter 8

Min Lu

Object:

Wald Confidence
interval
Profile Likelihood
Confidence Interval

R Example

Use the wchs data, choose your own outcome and predictors, and conduct a GLM with negative binomial family. Save the result in an excel (.csv) file.

**Class 6:
Chapter 8**

Min Lu

Object:

Wald Confidence
interval

Profile Likelihood
Confidence Interval

R Example

