Binomial Regression I

binomial regression with logit link

$$\log\left(\frac{p}{1-p}\right) = \eta = X\beta$$

"linear model of log odds"

Learning goals

Understand binomial regression

- know when you should use binormal regression.
- be able to write binomial regression model and its likelihood.
- be able to obtain estimators of parameters or function of parameters using R script.
- be able to quantify uncertainty of the estimators (e.g., computing CI).
- be able to test hypothesis.
- be able to do model selection.

Understand asymptotic properties of MLEs (maximum likelihood estimators)

- use asymptotic normality of MLEs to quantify uncertainty of the estimators (e.g., Wald CI).

Undertand Wald test and likelihood ratio test (LRT)

- use them to test hypothesis in binomial regression

Understand (scaled) deviance

- use it to test model adequacy or perform LRT and model comparison.

Challenger example

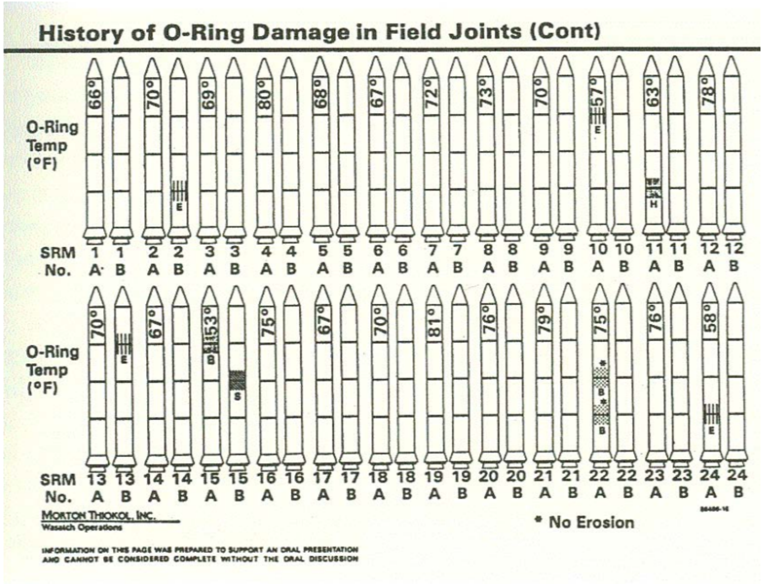
Challenger disaster

On the 28th of January 1986 the Space Shuttle Challenger broke apart after an O-ring seal failed at liftoff, leading to the deaths of its seven crew members.

Despite concerns about the O-rings failing due to the cold—the forecast temperature was $29 \pm 3\,^{o}F$—no one was able to provide an analysis that convinced NASA (who were under pressure to launch following several previous delays) not to go ahead.

The way the data was presented didn't help matters.

Challenger disaster: data



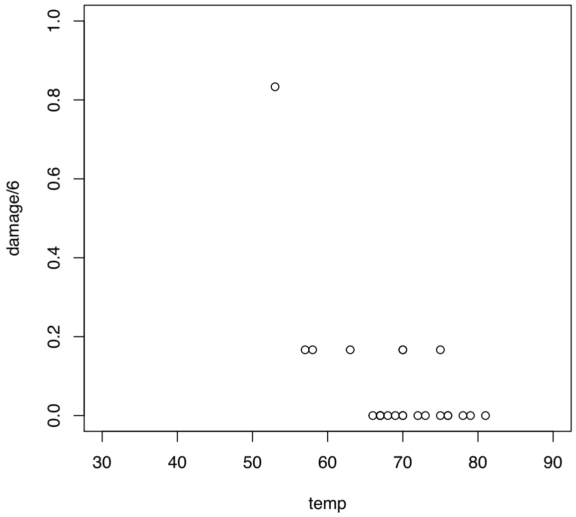
Challenger disaster: data

Summarise data using number of damaged O-rings (out of 6) per launch.

Data is in dataframe `orings`.
Response `damage` is number of damaged O-rings (out of 6).
Predictor `temp` is temperature (^oF)

```
> library(faraway)
> data(orings)
> str(orings)
data.frame: 23 obs. of  2 variables:
 $ temp  : num  53 57 58 63 66 67 67 67 68 69 ...
 $ damage: num  5 1 1 1 0 0 0 0 0 0 ...
> plot(damage/6 ~ temp, data = orings,
+      xlim = c(30, 90), ylim = c(0, 1))
```

Challenger disaster: visualize data



Challenger disaster: model

Make the assumption that Y_i, the number of damaged O-rings on the i-th launch, has distribution

$$Y_i \sim \text{bin}(6, p_i)$$

where p_i depends on the temperature t_i. We also assume that the Y_i are independent.

How to relate p_i with t_i.

- Question: just use a linear relationship? $p_i = \beta_0 + \beta_1 t_i$?

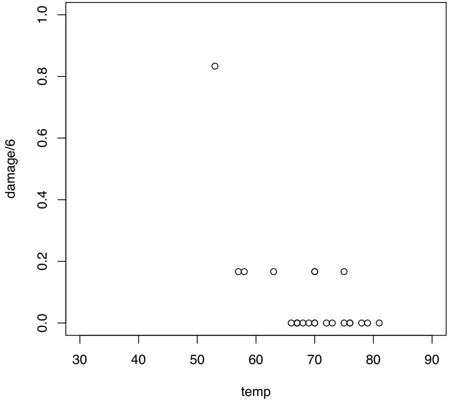
No. We need parameter bounds so $0 \leq p_i \leq 1$. We need alternate function to relate p_i with t_i.

Alternative function to relate p_i with t_i → for a single launch : use expectation

For a single launch, best estimate of p_i is just $y_i/6$. From plot of $y_i/6$ against t_i it is reasonable to assume that $p_i = p(t_i)$ where p is a smooth function of the temperature, decreasing from 1 down to 0 as the temperature increases.

$E[y_i] = np = 6p_i$

$E[p_i] = \frac{y_i}{6}$

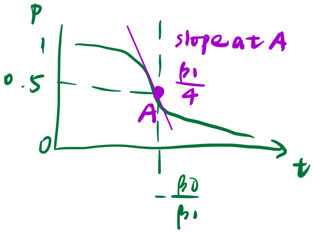


We choose logistic function: suppose that for some β_0 and β_1

$$p(t) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 t)}} = \frac{e^{\beta_0 + \beta_1 t}}{1 + e^{\beta_0 + \beta_1 t}}$$

- ① Restricts $0 \leq p(t) \leq 1$
- ② Monotonic increasing/decreasing function
- ③ Can be linearized through the *logit* transformation

$$\log \frac{p(t)}{1 - p(t)} = \beta_0 + \beta_1 t$$



We choose logistic function: suppose that for some β_0 and β_1

$$p(t) = \frac{1}{1 + e^{-(\beta_0+\beta_1 t)}} = \frac{e^{\beta_0+\beta_1 t}}{1 + e^{\beta_0+\beta_1 t}}$$

$p(t) = \frac{1}{2}$ when $\frac{e^0}{1+e^0}$ $\Rightarrow$ $\beta_0 + \beta_1 t = 0$ $t = -\frac{\beta_0}{\beta_1}$

- $p(t) = 1/2$ for $t = -\beta_0/\beta_1$, so $-\beta_0/\beta_1$ controls the location of the curve.
- $p'(-\beta_0/\beta_1) = \beta_1/4$, so β_1 controls the steepness of the curve.

$\beta_1 < 0$

Example: see R script and result in the section "logistic function with different values for beta0 and beta1" of Challenger.pdf.

Challenger disaster: model

Make the assumption that Y_i, the number of damaged O-rings on the i-th launch, has distribution

$$Y_i \sim \mathrm{bin}(6, p_i)$$

where

$$p_i = \frac{e^{\beta_0+\beta_1 t_i}}{1+e^{\beta_0+\beta_1 t_i}}.$$

parameter: β_0 & β_1

We also assume that the Y_i are independent.

Challenger disaster: model fitting

Make the assumption that Y_i, the number of damaged O-rings on the i-th launch, has distribution

$$Y_i \sim \text{bin}(6, p_i)$$

where

$$p_i = \frac{e^{\beta_0 + \beta_1 t_i}}{1 + e^{\beta_0 + \beta_1 t_i}}.$$

We also assume that the Y_i are independent.

Maximum likelihood estimators (MLE) $\hat{\beta}_0$, $\hat{\beta}_1$: values for β_0, β_1 which maximize the log-likelihood.

Challenger disaster: model fitting

The log-likelihood is

Y_i independent

$\log(AB)=\log A+\log B$

$$
\begin{aligned}
l(\beta_0,\beta_1) &= \log \mathcal{L}(\beta_0,\beta_1)\\
&= \log \mathbb{P}(\mathrm{Y}=\mathrm{y}\mid \beta_0,\beta_1) = \log \prod_i \mathbb{P}(Y_i=y_i\mid \beta_0,\beta_1)\\
&= \sum_i \log\left(\binom{6}{y_i} p_i^{y_i}(1-p_i)^{6-y_i}\right)\\
&= c+\sum_i \left(y_i\log p_i+(6-y_i)\log(1-p_i)\right)\\
&= c+\sum_i \left(y_i \boxed{\log\frac{p_i}{1-p_i}} + 6\log(1-p_i)\right)
\end{aligned}
$$

$\beta_0+\beta_1 t_i$

$$\log \frac{\frac{1}{1+e^{-\eta_i}}}{1-\frac{1}{1+e^{-\eta_i}}} = \log\frac{1}{e^{-\eta_i}} = \eta_i$$

$$\frac{1}{1-p_i} = 1+e^{\eta_i}$$

$$\log\left(\frac{1}{1-p_i}\right) = \log\left(1+e^{\eta_i}\right)$$

Put $\eta_i = \beta_0+\beta_1 t_i$, then $\log p_i/(1-p_i) = \eta_i$ and $\log(1-p_i) = -\log(1+e^{\eta_i})$.

$$\frac{p_i}{1-p_i} = e^{\eta_i} \qquad \frac{p_i-1+1}{1-p_i} = -1+\frac{1}{1-p_i} = 1+e^{\eta_i}$$

$$\log(1-p_i) = -\log(1+e^{\eta_i})$$

Challenger disaster: model fitting

There is no closed form solution for MLE $\hat{\beta}_0$, $\hat{\beta}_1$ in this model.
Numerical search procedures are required to find MLE.

- the `glm` function uses the iterative weighted least squares (IWLS) algorithm - I will cover this later.
- For now, let's use the `optim` function.

greedy method $\Rightarrow$ try different initial

Example:

- See R script and result in "maximum likelihood fitting" of Challenger.pdf
- $\hat{\beta}_0 = 11.667$ and $\hat{\beta}_1 =$ -0.216

Challenger disaster: questions

- Forecast probability of an O-ring being damaged when the launch temperature is 29 ^oF.
- How good is our forecast? Can we provide a confidence interval?
- Is temperature useful to predict the O-ring failing?

$\Downarrow$

may have other factors

linear model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2)$

$\hat{\beta_i} \quad \hat{\beta_i} \sim N(\beta_i, se(\hat{\beta_i}))$ 95% CI for $\beta_i = \hat{\beta_i} + Z_{\alpha/2} \cdot se(\hat{\beta_i})$

Binomial regression

Binomial regression model

Y_i independent

We suppose that we observe $Y_i \sim \text{bin}(m_i, p_i)$, $i = 1, \ldots, n$, independent.

The m_i are known and we suppose that for some **link function** g,

$$g(p_i) = \eta_i = \mathrm{x}_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_q x_{iq}$$

where x_i are known predictors and $\boldsymbol{\beta}$ are unknown parameters.

Binomial regression model: link function

Usual choices for g:

logit

$= x^T \beta$

$$\eta = \log \frac{p}{1-p}, \quad p = \underbrace{\frac{1}{1+\exp(-\eta)}}_{} = \frac{\exp(\eta)}{1+\exp(\eta)}$$

logistic function of predictor

complementary log-log

$$\eta = \log(-\log(1-p)), \quad p = 1 - \exp(-e^{\eta})$$

probit

$z \sim N(0,1)$

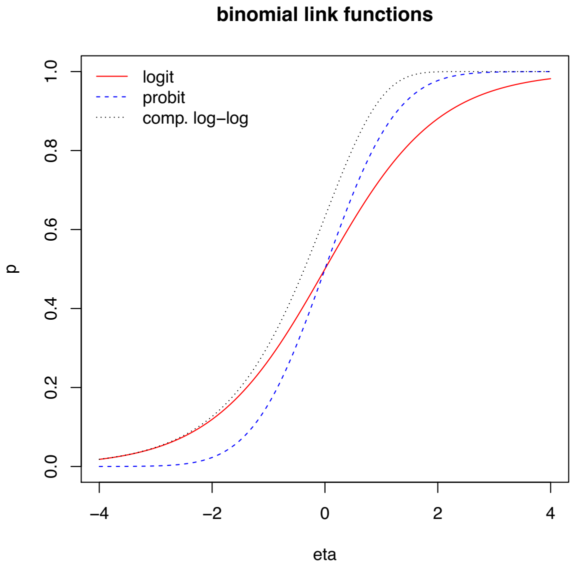
$$\eta = \Phi^{-1}(p), \quad p = \Phi(\eta), \quad p = P(z \le \eta)$$

where Φ is the cumulative distribution function (cdf) of the standard normal distribution.

Binomial regression model: link function

```
> curve(1/(1+exp(-x)), -4, 4, ylim=c(0,1),
+       xlab="eta", ylab="p", col="red",
+       main="binomial link functions")
> curve(pnorm(x), -4, 4, add=TRUE, col="blue", lty=2)
> curve(1-exp(-exp(x)), -4, 4, add=TRUE, col="black", lty=3)
> legend("topleft", c("logit", "probit", "comp. log-log"),
+        col=c("red", "blue", "black"), lty=c(1,2,3), bty="n")
```

Binomial regression model: link function



$g(p) = \eta$

$p = g^{-1}(\eta)$

Binomial regression model: likelihood

Given observations y_i of $Y_i \sim \text{bin}(m_i, p_i = g^{-1}(\eta_i))$, where $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$, the log-likelihood is

$$\begin{aligned} l(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \log \mathbb{P}(Y_i = y_i) \\ &= \sum_{i=1}^{n} \log \left(\binom{m_i}{y_i} p_i^{y_i} (1 - p_i)^{m_i - y_i} \right) \\ &= c + \sum_{i=1}^{n} y_i \log(g^{-1}(\eta_i)) + (m_i - y_i) \log(1 - g^{-1}(\eta_i)) \end{aligned}$$

There is no closed form solution for MLE of $\boldsymbol{\beta}$. Numerical search procedures are required to find MLE.

Reminder: questions from Challenger disaster

- Forecast probability of an O-ring being damaged when the launch temperature is $29\,^oF$.
- How good is our forecast? Can we provide a confidence interval?
- Is temperature useful to predict the O-ring failing?

Forecast probability of an O-ring being damaged when the launch temperature is 29 ^oF.

$\hat{p} = g^{-1}(\hat{\eta})$ when $t = 29$.

$\hat{p} = \frac{\exp(\hat{\eta})}{1+\exp(\hat{\eta})}$, where $\hat{\eta} = \hat{\beta}_0 + \hat{\beta}_1 29$

$\hat{p} = 0.995$: see R script and result in "prediction for temp of 29" of Challenger.pdf

Reminder: questions from Challenger disaster

- Forecast probability of an O-ring being damaged when the launch temperature is $29\,^oF$.
- How good is our forecast? Can we provide a confidence interval?
- Is temperature useful to predict the O-ring failing?

We need to know properties of MLE!!

Asymptotic properties MLE

Reminder: Maximum likelihood estimate (MLE)

Suppose that Y_i, $i = 1, \ldots, n$, are indepedendent, with densities/mass-functions $f_i(\cdot; \boldsymbol{\theta})$.

Given observations y_i of the Y_i, the log-likelihood is

$$l(\boldsymbol{\theta}) = l(\boldsymbol{\theta}; \mathrm{y}) = \sum_i \log f_i(y_i; \boldsymbol{\theta}).$$

MLE $\hat{\boldsymbol{\theta}}$ is that value of $\boldsymbol{\theta}$ which maximises $l(\boldsymbol{\theta})$.

Note that allowing f_i to depend on i means that we can include the case where the distribution of Y_i depends on some covariate x_i. That is, we can have $f_i(\cdot; \boldsymbol{\theta}) = f(\cdot; \mathrm{x}_i, \boldsymbol{\theta})$ for some common f.

Asymptotic properties of MLE

$n \to \infty$

Under certain regularity conditions (to be introduced later), the MLE is

- asymptotically consistent,
- asymptotically normal,
- asymptotically efficient.

MLE: asymptotic consistency

Let $\boldsymbol{\theta}^*$ denote a true value for $\boldsymbol{\theta}$. As $n \to \infty$, $\hat{\boldsymbol{\theta}} \xrightarrow{\mathrm{p}} \boldsymbol{\theta}^*$.

That is for any $\epsilon > 0$

$$\mathbb{P}(|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*| > \epsilon) \to 0 \text{ as } n \to \infty.$$

MLE: asymptotic normality

Let $\boldsymbol{\theta}^*$ denote a true value for $\boldsymbol{\theta}$.

$$\hat{\boldsymbol{\theta}} \approx^d N(\boldsymbol{\theta}^*, \boxed{\mathcal{I}(\boldsymbol{\theta}^*)^{-1}}).$$

MLE: asymptotic normality

The observed information is

$\nearrow$ 2nd partial derivative of log likelihood

$$\mathcal{J}(\boldsymbol{\theta}) = -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = -\mathrm{H}_{l(\boldsymbol{\theta})},$$

where the hessian matrix, $\mathrm{H}_{l(\boldsymbol{\theta})}$, is a square matrix of second-order partial derivatives of the log likelihood.

For binomial regression with one predictor (e.g., $\boldsymbol{\theta} = \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$):

$$\mathcal{J}(\boldsymbol{\theta}) = \mathcal{J}(\boldsymbol{\beta}) = \begin{pmatrix} -\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_0^2} & -\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_0} \\ -\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1} & -\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_1^2} \end{pmatrix}$$

Clearly $\mathcal{J}(\boldsymbol{\theta}) = \mathcal{J}(\boldsymbol{\theta}; \mathrm{y})$ depends on y through $l(\boldsymbol{\theta}) = l(\boldsymbol{\theta}; \mathrm{y})$.

MLE: asymptotic normality

The Fisher information is observed information

$$\mathcal{I}(\boldsymbol{\theta}) = \mathbb{E}\left[\boxed{\mathcal{J}(\boldsymbol{\theta}; \mathsf{Y})}\right].$$

$\mathcal{I}(\boldsymbol{\theta})$ does not depend on y. → since we did expectation

Exercise: binomial regression with a logit link

We suppose that we observe $Y_i \sim \text{bin}(m_i, p_i)$, $i = 1, \ldots, n$, independent, where $p_i = \frac{1}{1+\exp(-\eta_i)}$ and $\eta_i = \beta_0 + \beta_1 x_i$. The log-likelihood is

$$
\begin{aligned}
l(\beta_0, \beta_1) &= c + \sum_i \left[y_i \log \frac{p_i}{1-p_i} + m_i \log(1-p_i) \right] \\
&= c + \sum_i \left[y_i(\beta_0 + \beta_1 x_i) - m_i \log(1 + e^{\beta_0+\beta_1 x_i}) \right].
\end{aligned}
$$

$$
\frac{\partial l}{\partial \beta_0} = \sum_i y_i - \overbrace{m_i \frac{1}{1+e^{-\beta_0+\beta_1 x_i}}}^{m_i p_i} \qquad \frac{\partial l}{\partial \beta_0^2} = -\sum_i m_i p_i \frac{e^{\eta_i}}{(1+e^{\eta_i})^2}
$$

Exercise: binomial regression with a logit link

derivation !

Then,

$$\mathcal{J}(\boldsymbol{\beta}) = \begin{pmatrix} \sum_i m_i p_i(1-p_i) & \sum_i m_i x_i p_i(1-p_i) \\ \sum_i m_i x_i p_i(1-p_i) & \sum_i m_i x_i^2 p_i(1-p_i) \end{pmatrix}.$$

See my derivation in "observed_information_binomial_regression.pdf".

So, since there are no y_i terms left, → Expectation don't change

$$\mathcal{I}(\boldsymbol{\beta}) = \begin{pmatrix} \sum_i m_i p_i(1-p_i) & \sum_i m_i x_i p_i(1-p_i) \\ \sum_i m_i x_i p_i(1-p_i) & \sum_i m_i x_i^2 p_i(1-p_i) \end{pmatrix}.$$

MLE: asymptotic normality

Let $\boldsymbol{\theta}^*$ denote a true value for $\boldsymbol{\theta}$.

As $n \to \infty$,

$$\mathcal{I}(\boldsymbol{\theta}^*)^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \xrightarrow{\mathrm{d}} N(0, I).$$

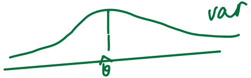
?

That is,

$$\hat{\boldsymbol{\theta}} \approx^d N(\boldsymbol{\theta}^*, \mathcal{I}(\boldsymbol{\theta}^*)^{-1}).$$

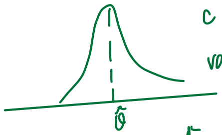
c small

var big

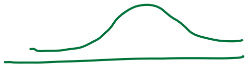


c big $\Rightarrow$ sharp $\Rightarrow$ more certain about estimator

var small



Any intuition for why variance is the inverse of fisher information which is related to hessian matrix?



$I(\theta) = E(-H)$

curvature $\uparrow$ H $\uparrow$ $I(\theta)\uparrow$ $\mathrm{var}(\hat{\theta})\downarrow$

MLE argmax $l(\theta)$

MLE: asymptotic efficiency

- Asymptotic consistency: $\hat{\boldsymbol{\theta}} \xrightarrow{\text{p}} \boldsymbol{\theta}^*$
- Asymptotic normality: $\hat{\boldsymbol{\theta}} \approx^d N(\boldsymbol{\theta}^*, \mathcal{I}(\boldsymbol{\theta}^*)^{-1})$

MLE: asymptotically unbiased estimator with smallest variance $\mathcal{I}(\boldsymbol{\theta}^*)^{-1}$.

Wald CI for $\boldsymbol{t}^T\boldsymbol{\theta}$ (linear combination of parameters)

use asymptotic normality

$t^T\theta = t_1\theta_1 + t_2\theta_2 + \cdots + t_m\theta_m$

$$\boldsymbol{t} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_m \end{pmatrix} \qquad \boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_m \end{pmatrix} \qquad \boldsymbol{\theta}^* = \begin{pmatrix} \theta_1^* \\ \theta_2^* \\ \vdots \\ \theta_m^* \end{pmatrix} \qquad \hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \vdots \\ \hat{\theta}_m \end{pmatrix}$$

$y = \beta_0 + \beta_1 x_i = [1 \quad x_i]\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$

$$\hat{\boldsymbol{\theta}} \approx N(\boldsymbol{\theta}^*, \mathcal{I}(\boldsymbol{\theta}^*)^{-1}).$$

In practice, we do not know $\boldsymbol{\theta}^*$, the true value of $\boldsymbol{\theta}$. Thus, we approximate $\mathcal{I}(\boldsymbol{\theta}^*)^{-1}$ using $\mathcal{I}(\hat{\boldsymbol{\theta}})^{-1}$, leading to

$$\hat{\boldsymbol{\theta}} \approx N(\boldsymbol{\theta}^*, \mathcal{I}(\hat{\boldsymbol{\theta}})^{-1}).$$

Then,

linear combination also follow normal distribution

$$\boldsymbol{t}^T\hat{\boldsymbol{\theta}} \approx N(\boldsymbol{t}^T\boldsymbol{\theta}^*, \boldsymbol{t}^T\mathcal{I}(\hat{\boldsymbol{\theta}})^{-1}\boldsymbol{t}).$$

Wald CI for $\boldsymbol{t}^T\boldsymbol{\theta}$ (linear combination of parameters)

$$\boldsymbol{t}^T\hat{\boldsymbol{\theta}} \approx N(\boldsymbol{t}^T\boldsymbol{\theta}^*, \boldsymbol{t}^T\mathcal{I}(\hat{\boldsymbol{\theta}})^{-1}\boldsymbol{t}).$$

An approximate $100(1-\alpha)\%$ confidence interval for $\boldsymbol{t}^T\boldsymbol{\theta}$:

$$\boldsymbol{t}^T\hat{\boldsymbol{\theta}} \pm z_\alpha\sqrt{\boldsymbol{t}^T\mathcal{I}(\hat{\boldsymbol{\theta}})^{-1}\boldsymbol{t}}, \qquad \text{where } \Phi(z_\alpha) = 1-\alpha/2$$

In particular, taking $\boldsymbol{t} =$ standard unit vectors, $\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}, \begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix}, \ldots, \begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}$,

we can obtain CI for each of parameters $\theta_1, \theta_2, \ldots, \theta_n$.
An approximate $100(1-\alpha)\%$ confidence interval for θ_i:

$I(\theta) = E[J(\theta, Y)]$ but when expectation is hard to compute or don't exist

$$\hat{\theta}_i \pm z_\alpha\sqrt{(\mathcal{I}(\hat{\boldsymbol{\theta}})^{-1})_{i,i}}$$

If $\mathcal{I}$ is unavailable then we can approximate it using the observed information $\mathcal{J}$, i.e., $\mathcal{I}(\hat{\boldsymbol{\theta}})^{-1} \approx \mathcal{J}(\hat{\boldsymbol{\theta}}; y)^{-1}$.

Reminder: questions from Challenger disaster

- Forecast probability of an O-ring being damaged when the launch temperature is 29 ^oF.
- How good is our forecast? Can we provide a confidence interval?
- Is temperature useful to predict the O-ring failing?

How good is our forecast? Can we provide a confidence interval?

not linearly combination of η

CI for $p = g^{-1}(\eta) = \frac{\exp(\eta)}{1+\exp(\eta)}$, where $\eta = \beta_0 + \beta_1 29$.

- Step 1: Compute a CI for η, (η_l, η_r).
- Step 2: CI for $p = g^{-1}(\eta)$ is $(g^{-1}(\eta_l), g^{-1}(\eta_r))$. since $g^{-1}(\eta)$ is monotonic increasing function

CI for p: (0.864307, 0.9998686). See R script and result in "Confidence Interval for p" of Challenger.pdf

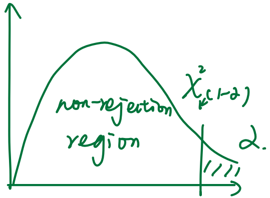
log likelihood ratio CI

$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

$X \sim N$

$\frac{x-\mu}{\sigma} \sim$ standard Normal

(standard Normal$)^2 \sim \chi^2$



$\hat{\theta} \sim N(\theta, se(\hat{\theta})^2)$
when $\theta = \theta^*$

$\frac{\hat{\theta}-\theta}{se(\hat{\theta})} \sim N(0,1)$

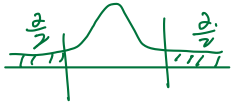
We have

$$2l(\hat{\boldsymbol{\theta}}) - 2l(\boldsymbol{\theta}^*) \approx \chi^2_k$$

where k is the dimension of $\boldsymbol{\theta}^*$.

This result can also, in principle, be used to construct a $100(1-\alpha)\%$ confidence region for $\boldsymbol{\theta}$:

compute non-rejection region



$$\{\boldsymbol{\theta} : 2l(\hat{\boldsymbol{\theta}}) - 2l(\boldsymbol{\theta}) \le \chi^2_k(1-\alpha)\}$$

wald CI use 2 approximation

① normality

② $I(\theta^*)^{-1} \to I(\hat{\theta})^{-1}$

where $\chi^2_k(1-\alpha)$ is the $100(1-\alpha)\%$ point for a χ^2_k distribution.

find θ fall in non-rejection region

One lab problem will be to plot CI using the log likelihood ratio.

log likelihood ration CI use 1 approximat.

$2l(\hat{\theta}) - 2l(\theta^*) \approx \chi^2_k$

$\left|\frac{\hat{\theta}-\theta}{se(\hat{\theta})}\right| \le z_{\frac{\alpha}{2}}$

$\Rightarrow \hat{\theta} \pm z_{\frac{\alpha}{2}} se(\hat{\theta})$

This approximation is generally better than the normal approximation for $\hat{\boldsymbol{\theta}}$. That is, it holds for smaller sample sizes.

MLE: regularity conditions

For maximum likelihood theory to hold we require

log-likelihood

- l smooth enough with respect to $\boldsymbol{\theta}$ (third derivatives exist and continuous)
- Third order derivatives of l have bounded expectations
- Support of Y_i does not depend on $\boldsymbol{\theta}$
- The domain Θ of $\boldsymbol{\theta}$ is finite dimensional and doesn't depend on Y_i
- $\boldsymbol{\theta}^*$ is not on the boundary of Θ.

References

- McCullagh & Nelder (1989), Appendix A.
- F.W. Scholz, Maximum likelihood estimation. *Encyclopedia of Statistical Sciences* Vol. 7, p.4629ff. Wiley, 2006.