Lab Session 1: Logistic regression

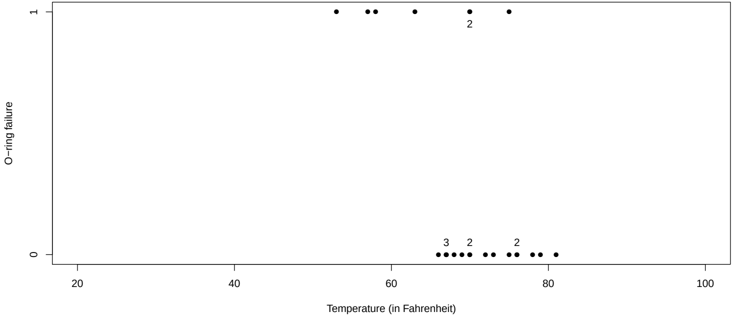
Hedibert Freitas Lopes

The University of Chicago Booth School of Business
5807 South Woodlawn Avenue, Chicago, IL 60637
http://faculty.chicagobooth.edu/hedibert.lopes

hlopes@ChicagoBooth.edu

O-Ring Data

Field O-ring failures ($y_i = 1$) in the 23 pre-*Challenger* space shuttle launches. *Challenger* was the shuttle that blew up on take off when the temperature (t_i) was 31 degrees Fahrenheit[1].



$Pr(\tilde{y} = 1|\tilde{t})$, for $\tilde{t}$ in $(20, 100)$?

[1] Christensen (1997) Log-Linear Models and Logistic Regression (2nd edition), pages 54-56.

Bernoulli model

For $i = 1, \ldots, n = 23$.

$$y_i|\theta_i \sim Bern(\theta_i)$$

Logit link:

$$\log\left(\frac{\theta_i}{1-\theta_i}\right) = \beta_0 + \beta_1(t_i - \bar{t}) = x_i'\beta$$

where $x_i' = (1, t_i - \bar{t})$, $\beta = (\beta_0, \beta_1)$ and $\bar{t} = 70$.

Likelihood for $\theta = (\theta_1, \ldots, \theta_n)$

$$l(\theta; y) = \prod_{i=1}^{n} \theta_i^{y_i}(1-\theta_i)^{1-y_i} = \left(\frac{\theta_i}{1-\theta_i}\right)^{y_i}(1-\theta_i)$$

It is easy to see that

$$\frac{\theta_i}{1-\theta_i} = e^{x_i'\beta} \quad \text{and} \quad (1-\theta_i) = \left(1 + e^{x_i'\beta}\right)^{-1}$$

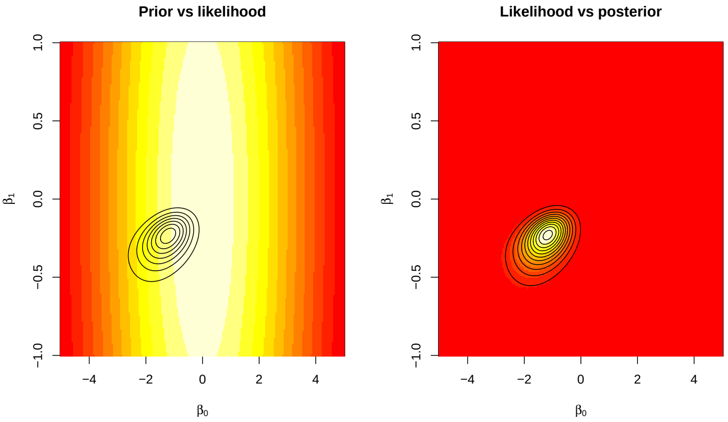
Likelihood of β is

$$L(\beta; y, x) = \prod_{i=1}^{n} \frac{e^{(x_i'\beta)y_i}}{1 + e^{x_i'\beta}}$$

Prior of β is $N(m_0, C_0)$ for

$$m_0 = 0_2 \quad \text{and} \quad C = I_2.$$

Prior, likelihood, posterior



Posterior and prior predictive

Posterior of β

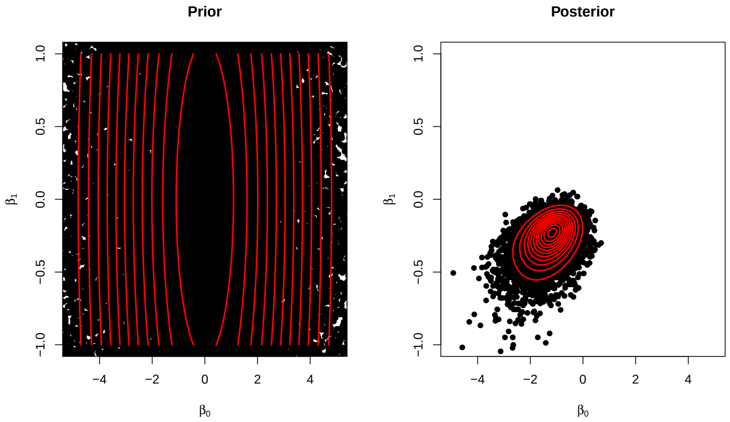
$$
\begin{aligned}
p(\beta|y,x) &= \frac{p(\beta)L(\beta;y,x)}{p(y|x)} \\
&\propto \exp\{-0.5(\beta - m_0)C_0^{-1}(\beta - m_0)\} \\
&\times \prod_{i=1}^{n} e^{(x_i'\beta)y_i}\left(1 + e^{x_i'\beta}\right)^{-1}.
\end{aligned}
$$

Prior predictive

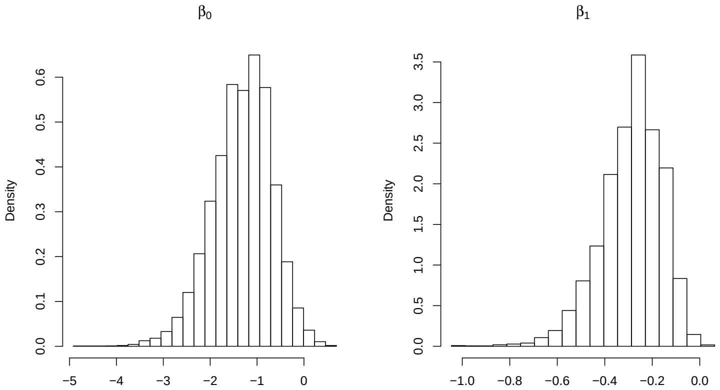
$$
\begin{aligned}
p(y|x) &= \int_0^1 L(\beta;y,x)p(\beta)d\beta \\
&= \int_0^1 \left[\prod_{i=1}^{n} e^{(x_i'\beta)y_i}\left(1 + e^{x_i'\beta}\right)^{-1}\right] p(\beta)d\beta.
\end{aligned}
$$

Sampling from $p(\beta|y)$ by SIR

Sample $\beta^{(1)}, \ldots, \beta^{(M)}$ from $p(\beta)$;
Resample from $\{\beta^{(1)}, \ldots, \beta^{(M)}\}$ with weights $L(\beta^{(i)}; y, x)$.



Marginal posterior



Parameter: mean (median), st dev, 95% interval

β_0: -1.3199 (-1.2657) 0.6333 (-2.6671 -0.1853)

β_1: -0.2880 (-0.2728) 0.1287 (-0.5774 -0.0772)

Posterior predictive

The probability of failure when the temperature is $\tilde{t}$ is

$$\begin{aligned} Pr(\tilde{y} = 1|\tilde{t}) &= E_{\beta|y,x}\left[Pr(\tilde{y} = 1|\tilde{t}, \beta)\right] \\ &= E_{\beta|y,x}\left(\frac{e^{\tilde{x}'\beta}}{1 + e^{\tilde{x}'\beta}}\right), \end{aligned}$$

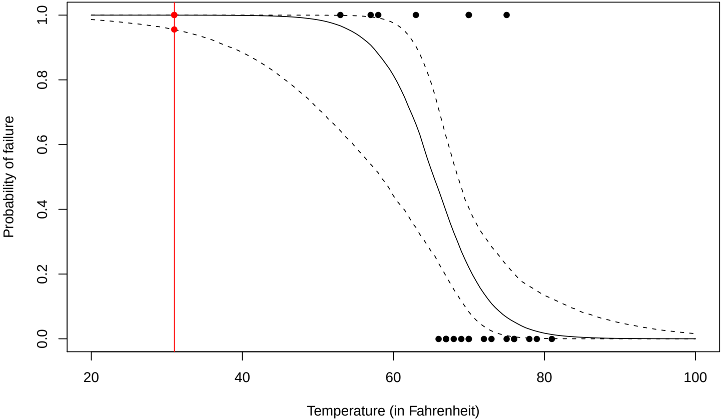
where $\tilde{x} = (1, \tilde{t} - \overline{t})'$

Approximation via MC integration

$$\hat{P}r(\tilde{y} = 1|\tilde{t}) = \frac{1}{M}\sum_{i=1}^{M}\left(\frac{e^{\tilde{x}'\beta^{(i)}}}{1 + e^{\tilde{x}'\beta^{(i)}}}\right)$$

where $\beta^{(1)}, \ldots, \beta^{(M)}$ are draws from $p(\beta|y, x)$.

$Pr(\tilde{y} = 1|\tilde{t} = 31)$?



95% posterior credibility interval for $Pr(\tilde{y} = 1|\tilde{t} = 31)$ is $(0.87, 100)$, with a median of 0.9998734.