3-17-15

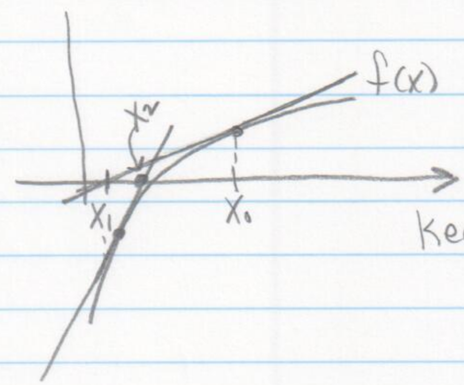
1 HW9

Newton-Raphson Method (or Newton's method)

* Used to find the root of a function

If f is your function of interest, use N-R to find a value r such that $f(r)=0$.

<u>Iterative Process</u>:
Choose initial x value x_0



Keep doing this until xs converge

The $(i+1)^{th}$ x-value, denoted by x_{i+1}, is found as the point at which the line tangent to f at x_i crosses the x-axis.

$x_{i+1} = ?$

Eqn for tanget line to f at x_i: $y = f'(x_i)x + b$

slope $= f'(x_i)$

intercept $= f(x_i) - f'(x_i)\cdot x_i = b$

$y = f'(x_i)x + (f(x_i) - f'(x_i)\cdot x_i)$

This line cuts through x-axis when $y=0$

$f'(x_i)x + (f(x_i) - f'(x_i)\cdot x_i) = 0$

$x = \frac{\cancel{f'(x_i)}x_i}{\cancel{f'(x_i)}} - \frac{f(x_i)}{f'(x_i)}$ $\quad x = x_i - \frac{f(x_i)}{f'(x_i)}$

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

*does not work when min or max value crosses x-axis

Use this iterative formula repeatedly unil the separation between consecutive x-values is sufficiently small.

Logistic Regression

Models a dichotomous dependent variable Y as a function of one or more independent variables X.

Let Y_i be the $\{0, 1\}$ indicator for individual i. We model Y_i as follows: $Y_i \sim Bern(p_i)$

$P(Y_i = 1) = p_i$

$P(Y_i = 0) = 1 - p_i = q_i$

For n independent Bernoulli r.v.s $Y_1, Y_2, Y_3, \ldots, Y_n$, each with their own Bernoulli distribution (with parameter p_i for corresponding Y_i), the likelihood function for the p_is is

$$L(p_1, p_2, p_3, \ldots, p_n) = (p_1)^{y_1}(1-p_1)^{1-y_1} \cdot (p_2)^{y_2}(1-p_2)^{1-y_2} \cdot \ldots (p_n)^{y_n}(1-p_n)^{1-y_n}$$

Now let p_i be a function of some independent variable x:

$p_i = \alpha + \beta x_i$ ← range of a line extends above 1 and below 0 so this is not a good model for p_i range exceeds $[0, 1]$

$$p_i = \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}$$ ← cannot be negative

← denom always > numerator so bounded by [0, 1]

Logistic model for p_i is a good choice, range is [0,1]

The Likelihood function is now dealing with only two parameters:

$$L(\alpha,\beta) =$$

$$\left(\frac{e^{\alpha+\beta x_1}}{1+e^{\alpha+\beta x_1}}\right)^{y_1}\left(\frac{1}{1+e^{\alpha+\beta x_1}}\right)^{1-y_1}\left(\frac{e^{\alpha+\beta x_2}}{1+e^{\alpha+\beta x_2}}\right)^{y_2}\left(\frac{1}{1+e^{\alpha+\beta x_2}}\right)^{1-y_2}\cdot \ldots$$

$$\left(\frac{e^{\alpha+\beta x_n}}{1+e^{\alpha+\beta x_n}}\right)^{y_n}\left(\frac{1}{1+e^{\alpha+\beta x_n}}\right)^{1-y_n}$$

Notice it is easier to work with $\ln(L)$

$$\ln(L(\alpha,\beta)) = \sum_{i=1}^{n} y_i\left[\left(\ln(e^{\alpha+\beta x_i}) - \ln(1+e^{\alpha+\beta x_i})\right)\right] +$$

$$(1-y_i)\left[\left(\ln(1) - \ln(1+e^{\alpha+\beta x_i})\right)\right]$$

$$= \sum_{i=1}^{n} y_i(\alpha+\beta x_i) \cancel{- y_i(\ln(1+e^{\alpha+\beta x_i}))} + \cancel{y_i \ln(1+e^{\alpha+\beta x_i})} - \ln(1+e^{\alpha+\beta x_i})]$$

$$= \sum_{i=1}^{n}\left[y_i(\alpha+\beta x_i) - \ln(1+e^{\alpha+\beta x_i})\right]$$ Natural log of Likelihood function

Joint posterior for α, β:

$$\pi_1(\alpha,\beta \mid \vec{y}) \propto \underbrace{L(\alpha,\beta)}_{e^{\Sigma(y_i(\alpha+\beta x_i) - \ln(1+e^{\alpha+\beta x_i})]}} \cdot \underbrace{\pi_0(\alpha,\beta)}_{\cdot\,(\text{joint prior for } \alpha,\beta)}$$

condition on data

Sample from the marginal posteriors for α and β using Metropolis (Hastings) Sampling:

1. Propose α^* from some proposal density
2. Accept α^* as a draw from $\pi_1(\alpha \mid \vec{y})$ with probability $\dfrac{\pi_1(\alpha^*, \beta \mid \vec{y})}{\pi_1(\alpha_c, \beta \mid \vec{y})}$

3. Propose β^* from some proposal density
4. Accept β^* as a draw from $\pi_1(\beta \mid \vec{y})$ with probability $\dfrac{\pi_1(\alpha, \beta^* \mid \vec{y})}{\pi_1(\alpha, \beta_c \mid \vec{y})}$

5. Repeat 1-4 many times

CPMA 573 — Homework #9

Exercise 1: Univariate Newton-Raphson (Newton's Method). Let $f(x) = x^m - c$ for integers $m > 1$ and $c > 0$. Show Newton's method is defined by

$$x_i = x_{i-1}\left(1 - \frac{1}{m} + \frac{c}{m x_{i-1}^m}\right),$$

and then find the root of $f(x) = x^3 - 5$. Compare your result with $\sqrt[3]{5}$.

Exercise 2: Metropolis-Hastings sampling and logistic regression. The first column of the data below show the number of days of radiotherapy received by each of 24 patients. The second column represents the absence (0) or presence (1) of disease at a site three years after treatment. A problem of interest is to use the covariate (days) to predict the disease presence at three years.

Days(X)	Response(Y)	Days(X)	Response(Y)
21	0	51	0
24	0	55	0
25	0	25	1
26	0	29	1
28	0	43	1
31	0	44	1
33	0	46	1
34	0	46	1
35	0	51	1
37	0	55	1
43	0	56	1
49	0	58	1

Model the Y_i's as independent Bernoulli random variables with disease-presence probability p_i, where

$$log\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta X_i,$$

and make Bayesian inference on α and β as follows:

- Write down the joint posterior distribution $\pi(\alpha, \beta | (\vec{x}, \vec{y}))$ for α and β. Use independent, normal priors for α and β with zero mean and variance 100.
- Use Metropolis-Hastings (with normal proposal densities) to sample 5000 independent (α, β) realizations.

- Provide trace-plots of the marginal posterior realizations for α and β, as well as accompanying 95% credible intervals for these two parameters.

Finally, produce a scatterplot of the (x, y) data, and superpose the posterior mean curve and credible bounds as follows:

- Discretize the x-axis into 250 points using `x <- seq(min(days),max(days),length=250)`
- Create a 5000×250 matrix, where the jth row corresponds to the evaluation of p at each of the 250 discretized x-values:

$$p = \frac{e^{\alpha_j + \beta_j x}}{1 + e^{\alpha_j + \beta_j x}}$$

- Use the `apply` function to find the mean of each of the 250 columns, the 97.5th percentile of each column, and the 2.5th percentile of each column. Superpose these on your x-y scatterplot.

plot data

m[1, mean]

plot means against x (250 xs)

Lisa Over

2. Likelihood $e^{\Sigma(y_i(\alpha+\beta x_i)-\ln(1+e^{\alpha+\beta x_i}))}$

Priors $N(0,100)$

$$\pi_0(\alpha)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(\alpha-\mu)^2/2\sigma^2}=\frac{1}{\sqrt{200\pi}}e^{-\alpha^2/200}$$

$$\pi_0(\beta)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(\beta-\mu)^2/2\sigma^2}=\frac{1}{\sqrt{200\pi}}e^{-\beta^2/200}$$

$\pi(\alpha,\beta|\vec{x},\vec{y})=$

$$\frac{1}{200\pi}\cdot e^{\Sigma(y_i(\alpha+\beta x_i)-\ln(1+e^{\alpha+\beta x_i}))}\cdot e^{-\alpha^2/200}\cdot e^{-\beta^2/200}$$

$$=\frac{1}{200\pi}e^{[\Sigma y_i\alpha+\Sigma y_i\beta x_i-\Sigma\ln(1+e^{\alpha+\beta x_i}]-\alpha^2/200-\beta^2/200}$$ ← Joint Posterior

Full Conditional

$$\pi_1(\alpha|\vec{x},\vec{y})\propto e^{\Sigma y_i\alpha-\Sigma\ln(1+e^{\alpha+\beta x_i})-\alpha^2/200}$$

$$\pi_1(\beta|\vec{x},\vec{y})\propto e^{\Sigma y_i\beta x_i-\Sigma\ln(1+e^{\alpha+\beta x_i})-\beta^2/200}$$

target ratio for α

$$e^{[\Sigma y_i\alpha^*-\Sigma\ln(1+e^{\alpha^*+\beta x_i})-\alpha^{*2}/200]-[\Sigma y_i\alpha_c-\Sigma\ln(1+e^{\alpha_c+\beta x_i})-\alpha_c^2/200]}$$

target ratio for β

$$e^{[\Sigma y_i\beta^* x_i-\Sigma\ln(1+e^{\alpha+\beta^* x_i})-\beta^{*2}/200]-[\Sigma y_i\beta_c x_i-\Sigma\ln(1+e^{\alpha+\beta_c x_i})-\beta_c^2/200]}$$