36-711 – FingerEx 09 Solutions

1

(a) Since independent, log-likelihood is $\ell_n(\theta) = log(\prod_{i=1}^n (\frac{1}{1+exp[-\mathbf{x}_i^T\theta]})^{y_i}(\frac{exp[-\mathbf{x}_i^T\theta]}{1+exp[-\mathbf{x}_i^T\theta]})^{1-y_i}) = \sum_{i=1}^n (1-y_i)log(exp[-\mathbf{x}_i^T\theta]) - log(1+exp[-\mathbf{x}_i^T\theta])$, therefore $\nabla\ell_n(\theta) = \sum_{i=1}^n[(1-y_i)(-\mathbf{x}_i) + \frac{exp[-\mathbf{x}_i^T\theta]}{1+exp[-\mathbf{x}_i^T\theta]}(-\mathbf{x}_i)] = \sum_{i=1}^n[(y_i - \frac{1}{1+exp[-\mathbf{x}_i^T\theta]})\mathbf{x}_i] = \mathbf{X}^T[\mathbf{Y} - \mu(\theta)]$

(b) Negative Hessian $-\mathbf{H}_n(\theta) = -\nabla^2\ell_n(\theta) = \nabla[\mathbf{X}^T\mu(\theta)] = \sum_{i=1}^n[\frac{-exp[-\mathbf{x}_i^T\theta]}{(1+exp[-\mathbf{x}_i^T\theta])^2}(-\mathbf{x}_i)\mathbf{x}_i^T] = \sum_{i=1}^n[p_i(1-p_i)\mathbf{x}_i\mathbf{x}_i^T] = \mathbf{X}^T\Sigma\mathbf{X}$. Notice this actually doesn't depend on $\mathbf{Y}$, therefore $\mathbf{J}_n(\theta) = E(-\nabla^2\ell_n(\theta)) = \mathbf{X}^T\Sigma\mathbf{X}$ too.

2

(b)
```
Y_c(rep(0,55),rep(1,2),rep(0,155),rep(1,7),rep(0,152),rep(1,3),rep(0,13))
x1_rep(1,length(Y))
x2_c(rep(7,55),rep(14,157),rep(27,159),rep(57,16))
COX_data.frame(cbind(Y, x1, x2))

## single term in the negative log-likelihood
nllterm_formula( ~log(1+exp(- theta1*x1 - theta2*x2))
                 - (1-Y)*log(exp(- theta1*x1 - theta2*x2)) )

library(MASS)

## obtain gradient and hessian
nll3_deriv3(nllterm, c("theta1","theta2"), function(theta1, theta2) NULL)
ans_ms(~nll3(theta1, theta2), data=COX, start=c(theta1=1, theta2=1), trace=T)

## Estimates
>summary(ans)
Final value: 47.89436

Solution:
              Par.         Grad. Hessian.tht1 Hessian.tht2
theta1 -5.13246849 3.587131e-013     10.99565    338.49550
theta2  0.06769814 5.499601e-012    338.49550  13319.52578

Information:
            tht1          tht2
tht1  0.41783204 -0.0106185663
tht2 -0.01061857  0.0003449325

Convergence: BOTH X- AND RELATIVE FUNCTION CONVERGENCE

Computations done:
 Iterations Function Gradient
         13       17       13

## Standard Errors
> sqrt(diag(summary(ans)$Information))
[1] 0.64639929 0.01857236
```

(c)
```
## GLM yields the same result as ms()
ans2_glm(Y ~ x1+x2-1, family=binomial, data=COX)
> summary(ans2)[3]
$coefficients:
        Value Std. Error   t value
x1 -5.1324663 0.64596781 -7.945390
x2  0.0676981 0.01856673  3.646205
```

3

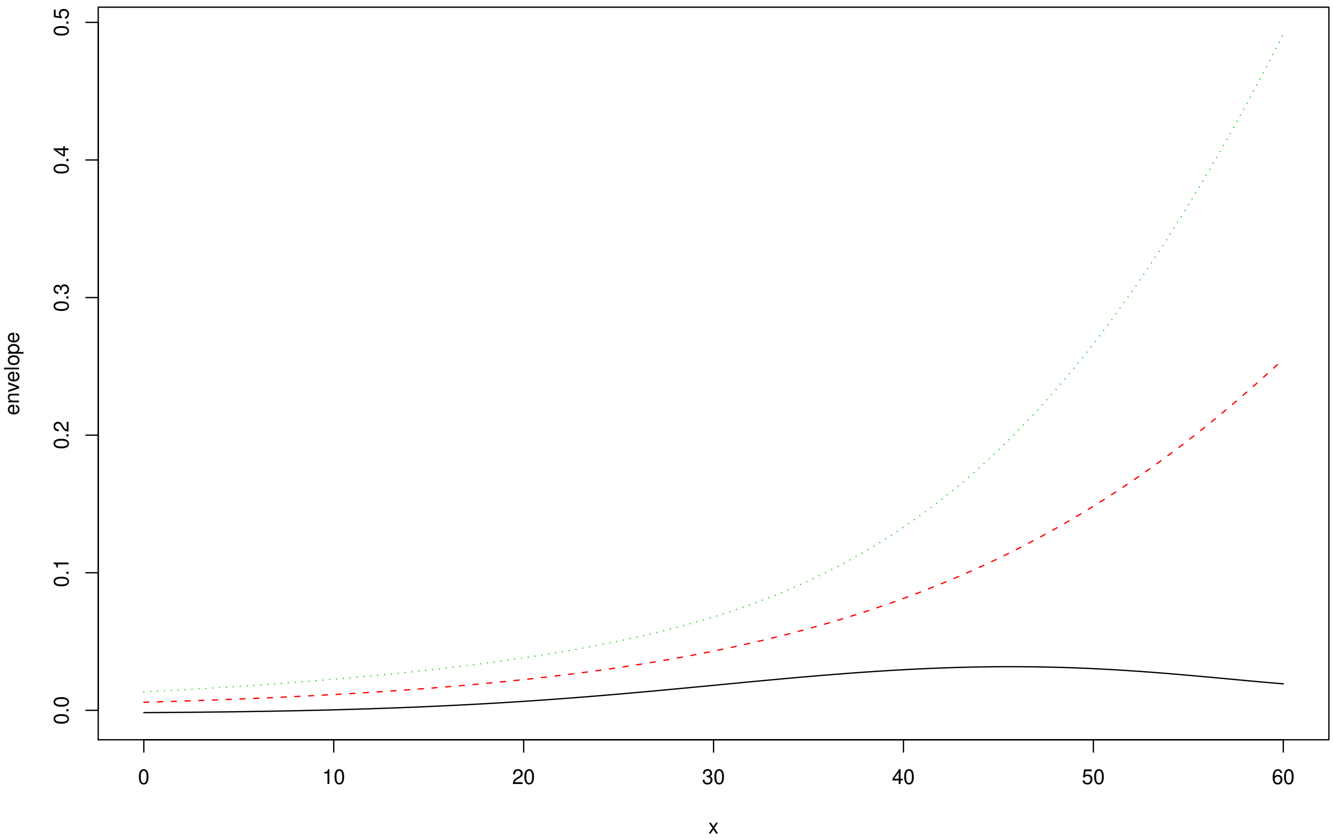
(a)
```
## obtain the estimate and graident of G(theta)
fcn_deriv3( ~ 1/(1 + exp(-theta1*x1 - theta2*x2)), c("theta1", "theta2"),
              function(theta1, theta2, x1, x2) NULL)
estimate_as.numeric(fcn(ans$parameters[1], ans$parameters[2],1 , 42))
grad_attr(fcn(ans$parameters[1], ans$parameters[2], 1, 42), "gradient")

> grad
        theta1   theta2
[1,] 0.0835571 3.509397

## calculate variance using delta method
v_grad %*% summary(ans)$Information %*% t(grad)
> v
            [,1]
[1,] 0.000937886

## calculate approx 95% C.I. of G(theta)
> estimate + c(-2, 0, 2) * sqrt(v)
[1] 0.03077599 0.09202583 0.15327568
```

(b)
```
x_seq(0,60,by=0.01)
estimates_as.numeric(fcn(ans$parameters[1], ans$parameters[2], 1, x))
grads_attr(fcn(ans$parameters[1], ans$parameters[2], 1, x), "gradient")
vars_grads %*% summary(ans)$Information %*% t(grads)
SEs_sqrt(diag(vars))
envelope_estimates + matrix(SEs, ncol=1) %*% matrix(c(-2,0,2), nrow=1)
matplot(x, envelope, type="l")
```



4

(a) This is because $E_\theta[\nabla \ell_n(\theta)] = 0$. Assumptions are needed to permit interchange of derivative operation and integral. (see Monahan p203 for details)

(b) First notice $E_\theta[\nabla \ell_n(\theta)] = 0 = E_\theta[\nabla B(\theta) + K(Y)^T \nabla\gamma(\theta)] = \nabla B(\theta) + \mu(\theta)^T \nabla\gamma(\theta) \Longrightarrow \nabla B(\theta) = -\mu(\theta)^T \nabla\gamma(\theta)$. Assume integration and differentiation can be interchanged,

$$\begin{aligned}
\nabla\mu(\theta) &= \int K(Y) \nabla L(\theta) d\nu(Y) \\
&= \int K(Y) L(\theta)[\nabla B(\theta) + K(Y)^T \nabla\gamma(\theta)] d\nu(Y) \\
&= \int K(Y) L(\theta)[K(Y)^T - \mu(\theta)^T] \nabla\gamma(\theta) d\nu(Y) \\
&= \{\int K(Y) K(Y)^T L(\theta) d\nu(Y) - \int K(Y) \mu(\theta)^T L(\theta) d\nu(Y)\} \nabla\gamma(\theta) \\
&= \Sigma(\theta) \nabla\gamma(\theta)
\end{aligned}$$