240D Winter 2011 Solutions to Midterm Exam

1.(a) Here

$$\begin{aligned} Q_N(\boldsymbol{\beta}) &= N^{-1}\sum_i(y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}))^2 \\ &= N^{-1}\sum_i\{[y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)]+[\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)-\exp(\mathbf{x}_i'\boldsymbol{\beta})]\}^2 \\ &= N^{-1}\sum_i\{[y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)]^2+2[y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)][\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)-\exp(\mathbf{x}_i'\boldsymbol{\beta})] \\ &\quad +[\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)-\exp(\mathbf{x}_i'\boldsymbol{\beta})]^2\}. \end{aligned}$$

(b) Now

$$\begin{aligned} Q_0(\boldsymbol{\beta}) &= \operatorname{plim} Q_N(\boldsymbol{\beta}) \\ &= \operatorname{plim} N^{-1}\sum_i[y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)]^2 \\ &\quad +2\operatorname{plim} N^{-1}\sum_i[y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)][\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)-\exp(\mathbf{x}_i'\boldsymbol{\beta})] \\ &\quad +\operatorname{plim} N^{-1}\sum_i[\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)-\exp(\mathbf{x}_i'\boldsymbol{\beta})]^2 \\ &= \lim N^{-1}\sum_i \mathrm{E}_{\mathbf{x}i}[\sigma_{i0}^2]+0+\lim N^{-1}\sum_i \mathrm{E}_{\mathbf{x}_i}[[\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)-\exp(\mathbf{x}_i'\boldsymbol{\beta})]]^2 \end{aligned}$$

where apply a LLN to first term and use $\mathrm{E}[[y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)]^2]=\mathrm{E}_{\mathbf{x}_i}[[y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)]^2|\mathbf{x}_i]=\mathrm{E}_{\mathbf{x}_i}[\mathrm{V}[y_i|\mathbf{x}_i]]=\mathrm{E}_{\mathbf{x}_i}[\sigma_{i0}^2]$ where second last equality uses $\mathrm{E}[y_i|\mathbf{x}_i]=\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)$ and a LLN to the second term using $\mathrm{E}[y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)|\mathbf{x}_i]=0$ since $\mathrm{E}[y_i|\mathbf{x}_i]=\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)$.

(c) $\widehat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}_0$ since by inspection $\operatorname{plim} Q_N(\boldsymbol{\beta})$ is clearly minimized at $\boldsymbol{\beta}=\boldsymbol{\beta}_0$, as then the third term (the only term involving $\boldsymbol{\beta}$) takes the minimum value of zero.
Or if you want to do more work in this example: differentiate w.r.t. $\boldsymbol{\beta}$ (not $\boldsymbol{\beta}_0$)

$$\begin{aligned} \frac{\partial Q_0(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} &= \lim N^{-1}\sum_i -2\mathrm{E}_{\mathbf{x}_i}[\exp(\mathbf{x}_i'\boldsymbol{\beta})[\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)-\exp(\mathbf{x}_i'\boldsymbol{\beta})]] \\ &= 0 \text{ when } \boldsymbol{\beta}=\boldsymbol{\beta}_0. \end{aligned}$$

(d) Consider the first term: $\operatorname{plim}\frac{1}{N}\sum_i(y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2$.
This is the average $\overline{X}_N=N^{-1}\sum_i X_i$ where $X_i=(y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2$.
Here X_i is i.i.d. as $(\mathbf{x}_i,y_i)$ are iid - best to use Khinchines Theorem (as then minimal assumptions).
This requires $\mathrm{E}[X_i]=\mathrm{E}[(y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2]=\mathrm{E}_{\mathbf{x}i}[\sigma_{i0}^2]$ exists.
Similarly for the second and third terms.

(e) No. In (b) the second term will no longer disappear if $\mathrm{E}[y_i|\mathbf{x}_i]\neq\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)$, so $Q_0(\boldsymbol{\beta})$ will not necessarily be maximized at $\boldsymbol{\beta}=\boldsymbol{\beta}_0$. (Also the first term will change and now involve $\boldsymbol{\beta}$.)
So $\widehat{\boldsymbol{\beta}}$ will be consistent if the functional form for $\mathrm{E}[y_i|\mathbf{x}_i]$ is misspecified.

2.(a) Differentiating $Q_N(\boldsymbol{\beta})=\frac{1}{N}\sum_i(y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}))^2$

$$\begin{aligned} \frac{\partial Q_N}{\partial\boldsymbol{\beta}} &= \tfrac{1}{N}\sum_i -2(y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}))\exp(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i \\ &= \tfrac{-2}{N}\sum_i(y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}))\exp(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i. \end{aligned}$$

Can assume the result that $\sqrt{N}\left.\frac{\partial Q_N}{\partial\boldsymbol{\beta}}\right|_{\boldsymbol{\beta}_0}\xrightarrow{d}\mathcal{N}[\mathbf{0},\,\mathbf{B}_0]$, where

$$\begin{aligned} \mathbf{B}_0 &= \operatorname{plim} N\left.\frac{\partial Q_N}{\partial\boldsymbol{\beta}}\frac{\partial Q_N}{\partial\boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}_0} \\ &= \operatorname{plim} N\left(\tfrac{-2}{N}\sum_i(y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)\mathbf{x}_i\right)\left(\tfrac{-2}{N}\sum_i(y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)\mathbf{x}_i\right)' \\ &= \lim\tfrac{4}{N}\sum_i\mathrm{E}\left[(y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2(\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2\mathbf{x}_i\mathbf{x}_i'\right] \text{ using independence and } \mathrm{E}[y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)|\mathbf{x}_i]=0 \\ &= \lim\tfrac{4}{N}\sum_i\mathrm{E}_{\mathbf{x}_i}[\sigma_{i0}^2(\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2\mathbf{x}_i\mathbf{x}_i']. \end{aligned}$$

(b) We have differentiating $\partial Q_N/\partial\beta$ by parts

$$\begin{aligned}\frac{\partial^2 Q_N}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} &= \tfrac{2}{N}\textstyle\sum_i(\exp(\mathbf{x}_i'\boldsymbol{\beta}))^2\mathbf{x}_i\mathbf{x}_i' \;-\; \tfrac{2}{N}\textstyle\sum_i(y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}))(\exp(\mathbf{x}_i'\boldsymbol{\beta}))\mathbf{x}_i\mathbf{x}_i' \\ \mathbf{A}_0 &= \text{plim}\left.\frac{\partial^2 Q_N}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}_0} = \lim \tfrac{2}{N}\textstyle\sum_i(\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2\mathbf{x}_i\mathbf{x}_i', \quad \text{since } \mathrm{E}[y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)|\mathbf{x}_i]=0.\end{aligned}$$

(c) Combining and noting that $(\frac{-2}{N})^{-1}(\frac{4}{N})(\frac{-2}{N})^{-1} = (\frac{1}{N})^{-1}(\frac{1}{N})(\frac{1}{N})^{-1}$

$$\begin{aligned}&\sqrt{N}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \xrightarrow{d} \mathcal{N}[\mathbf{0},\ \mathbf{A}_0{}^{-1}\mathbf{B}_0\mathbf{A}_0{}^{-1}] \\ &\xrightarrow{d} \mathcal{N}[\mathbf{0},\ \left(\lim \tfrac{1}{N}\textstyle\sum_i \mathrm{E}[\sigma_{i0}^2(\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2\mathbf{x}_i\mathbf{x}_i']\right)^{-1}\left(\lim \tfrac{1}{N}\textstyle\sum_i\{\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)\}^2\mathbf{x}_i\mathbf{x}_i'\right)()^{-1}].\end{aligned}$$

(d) Use $\widehat{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{A}}^{-1}$ where $\widehat{\mathbf{A}} = \frac{1}{N}\sum_i \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}))^2(\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}))^2\mathbf{x}_i\mathbf{x}_i']$
and $\widehat{\mathbf{B}} = \frac{1}{N}\sum_i(y_i-\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}))^2(\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}))^2\mathbf{x}_i\mathbf{x}_i'$.

(e) Reject $H_0:\beta_j=0$ against $H_a:\beta_j\neq 0$ at level a if $|t|>1.96$
where $t=\widehat{\beta}_j/s_{\widehat{\beta}_j}$ where $s^2_{\widehat{\beta}_j}$ is the j^{th} diagonal entry in the matrix in (g).

3.(a) A sequence of random variables $\{b_N\}$ converges in probability to b if for any $\varepsilon>0$ and $\delta>0$, there exists $N^*=N^*(\varepsilon,\delta)$ such that for all $N>N^*$, $\Pr[|b_N-b|<\varepsilon]>1-\delta$.

(b) A sequence of random variables $\{b_N\}$ converges in distribution to a random variable b if $\lim_{N\to\infty} F_N = F$, at every continuity point of F, where F_N is the distribution of b_N, F is the distribution of b, and convergence is in the usual mathematical sense.

(c) In general it will not, unless b is a constant, rather than a random variable.

(d) Because the statistics we use, such as estimators, involve averages.

(e) Newton Raphson: $\widehat{\boldsymbol{\theta}}_{s+1}-\widehat{\boldsymbol{\theta}}_s = -[\partial^2 Q_N(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'|_{\widehat{\boldsymbol{\theta}}_s}]^{-1}\times\partial Q_N(\boldsymbol{\theta})/\partial\boldsymbol{\theta}|_{\widehat{\boldsymbol{\theta}}_s}$.

(f) When y_i equals 1 with probability $\Lambda(\mathbf{x}_i'\boldsymbol{\beta})$ and 0 with probability $1-\Lambda(\mathbf{x}_i'\boldsymbol{\beta})$,
we can write the density as $f(y_i|\mathbf{x}_i)=\Lambda(\mathbf{x}_i'\boldsymbol{\beta})^{y_i}(1-\Lambda(\mathbf{x}_i'\boldsymbol{\beta}))^{1-y_i}$.
Then $\ln L = \ln\left(\prod_{i=1}^N f(y_i|\mathbf{x}_i)\right) = \sum_{i=1}^N \ln f(y_i|\mathbf{x}_i) = \sum_{i=1}^N \ln(\Lambda(\mathbf{x}_i'\boldsymbol{\beta})^{y_i}(1-\Lambda(\mathbf{x}_i'\boldsymbol{\beta}))^{1-y_i})$
$= \sum_{i=1}^N\{y_i\ln\Lambda(\mathbf{x}_i'\boldsymbol{\beta})+(1-y_i)\ln(1-\Lambda(\mathbf{x}_i'\boldsymbol{\beta})).\}$

4. I expected answer to cover both parameter interpretation (ME's) and statistical significance.
The coefficients are jointly statistically significant at 5%.
Individually the coefficient of x_2 is statistically significant at 5% but that of x_3 is not.
Here $\mathrm{ME}_j = g'(\beta_1+\beta_2x_2+\beta_3x_3)\beta_j$ where $g'(\cdot)<0$ so sign of ME is the reverse of the sign of β_j.
It follows that $\mathrm{E}[y|x_2,x_3]$ decreases when x_2 increases and when x_3 increases.
As $.42/.27\simeq 1.5$, a one unit change in x_3 has about 1.5 times the effect of a one unit change in x_2.
Aside: Also since standard errors reported were not labelled "robust" might also want to check this.

Overall - question 1 done well. Questions 2-4 not done as well. Some fundamental errors were made at times - so check for my hand-written comments on your answers.
The course grade will be based on a curve from the combined scores of midterm (35%), final (50%) and assignments (15%).
The curve for this exam is only a guide to give you a rough idea of how you are doing.

Scores out of	35
$75th$ percentile	28 (83%)
Median	26.5 (76%)
$25th$ percentile	21.5 (61%)

A	28 and above
A-	22 and above
B+	17 and above