1. (a) Here
\[ \ln f(y) = \ln(y + 1) + y \ln 2 + y \ln \lambda - (y + 0.5) \ln(1 + 2\lambda) \text{ and } \lambda = \exp(x'\beta) \text{ and } \ln \lambda = x'\beta \]
\[ \Rightarrow Q_N(\beta) = \frac{1}{N} \sum_i \ln f(y_i) = \frac{1}{N} \sum_i \{ \ln(y_i + 1) + y_i \ln 2 + y_i x_i' \beta - (y_i + 0.5) \ln(1 + 2 \exp(x_i' \beta)) \}. \]

(b) Now
\[ Q_0(\beta) = \plim Q_N(\beta) \]
\[ = \plim \frac{1}{N} \sum_i \ln(y_i + 1) + \plim \frac{1}{N} \sum_i y_i \ln 2 + \plim \frac{1}{N} \sum_i y_i x_i' \beta \]
\[ - \plim \frac{1}{N} \sum_i (y_i + 0.5) \ln(1 + 2 \exp(x_i' \beta)) \]
\[ = \lim \frac{1}{N} \sum_i E[\ln(y_i + 1)] + \lim \frac{1}{N} \sum_i E[y_i \ln 2] + \lim \frac{1}{N} \sum_i E[y_i x_i' \beta] \]
\[ - \lim \frac{1}{N} \sum_i E[(y_i + 0.5) \ln(1 + 2 \exp(x_i' \beta))] \text{ if LLN can be applied} \]
\[ = \lim \frac{1}{N} \sum_i E[\ln(y_i + 1)] + \lim \frac{1}{N} \sum_i E[\exp(x_i' \beta_0) \ln 2 + \frac{1}{N} \sum_i \exp(x_i' \beta_0) x_i' \beta] \]
\[ - \lim \frac{1}{N} \sum_i (\exp(x_i' \beta_0) + 0.5) \ln(1 + 2 \exp(x_i' \beta)) \text{ as } E[y_i x_i] = \exp(x_i' \beta_0) \]

(c) Differentiate w.r.t. \( \beta \) (not \( \beta_0 \))
\[ \frac{\partial Q_0(\beta)}{\partial \beta} = \lim \frac{1}{N} \sum_i \exp(x_i' \beta_0) x_i - \lim \frac{1}{N} \sum_i (\exp(x_i' \beta_0) + 0.5) \frac{2 \exp(x_i' \beta)}{1 + 2 \exp(x_i' \beta)} x_i \]
\[ = \lim \frac{1}{N} \sum_i \exp(x_i' \beta_0) x_i - \lim \frac{1}{N} \sum_i (1 + 2 \exp(x_i' \beta_0)) \frac{\exp(x_i' \beta)}{1 + 2 \exp(x_i' \beta)} x_i \]
\[ = \lim \frac{1}{N} \sum_i \exp(x_i' \beta_0) x_i - \lim \frac{1}{N} \sum_i \exp(x_i' \beta_0) x_i \text{ when } \beta = \beta_0 \]
\[ = 0. \]

Since plim \( Q_N(\beta) \) attains a local maximum at \( \beta = \beta_0 \),
Conclude that \( \hat{\beta} \) that solve \( \partial Q_N(\beta)/\partial \beta \) is consistent for \( \beta_0 \).

[Question said no need to check that \( \partial^2 Q_0(\beta)/\partial \beta \partial \beta' \) at \( \beta_0 \) is negative definite.]

(d) Consider the third term: \( \plim \frac{1}{N} \sum_i y_i x_i' \beta \).
This is the average \( X_N = N^{-1} \sum_i X_i \) where \( X_i = y_i x_i' \beta \).
Here \( X_i = y_i x_i' / \beta \) is i.n.i.d. as \( x_i \) is nonstochastic. So need Markov LLN.
Note that if \( x_i \) is stochastic then we could have \( y_i x_i' \beta \) i.i.d. even though \( E[y_i x_i] \) varies with \( x_i \).
With \( \delta = 1 \) this requires that \( V[y_i | x_i] \) exists and \( x_i \) is bounded.
[More formally, with \( \delta = 1 \) we need \( \sum_{i=1}^\infty E[((y_i x_i' \beta - E[y_i x_i' \beta])^2)]/i = \sum_{i=1}^\infty V[y_i | x_i] (x_i' \beta)^2 / i < \infty \)]

(e) From (c) and (d) all we need for consistency is \( E[y_i | x_i] = \exp(x_i' \beta_0) \).
Provided this holds the MLE is consistent even if other aspects of the distribution are misspecified.
If \( E[y_i | x_i] \neq \exp(x_i' \beta_0) \) then the MLE is inconsistent.
2. (a) Differentiating \( S_N(\beta) = \frac{1}{2N} \sum_i (y_i - g(x'_i\beta))^2 \)

\[
\frac{\partial S_N}{\partial \beta} = \frac{1}{2N} \sum_i -2(y_i - g(x'_i\beta))g'(x'_i\beta)x_i = \frac{-1}{N} \sum_i (y_i - g(x'_i\beta))g'(x'_i\beta)x_i,
\]

where \( g'(z) = \partial g(z)/\partial z \).

Can assume the result that \( \sqrt{N} \frac{\partial S_N}{\partial \beta} \bigg|_{\beta_0} \xrightarrow{d} \mathcal{N}(0, B_0) \), where

\[
B_0 = \lim N \frac{\partial^2 S_N}{\partial \beta \partial \beta'} \bigg|_{\beta_0} = \lim \frac{1}{N} \sum_i g'(x'_i\beta)g'(x'_i\beta)x_ix'_i + \frac{1}{N} \sum_i (y_i - g(x'_i\beta))g''(x'_i\beta)x_ix'_i.
\]

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

\[
\frac{\partial^2 S_N}{\partial \beta \partial \beta'} = \frac{-1}{N} \sum_i g'(x'_i\beta)g'(x'_i\beta)x_ix'_i + \frac{1}{N} \sum_i (y_i - g(x'_i\beta))g''(x'_i\beta)x_ix'_i
\]

\[
A_0 = \lim \frac{\partial^2 S_N}{\partial \beta \partial \beta'} \bigg|_{\beta_0} = -\lim \frac{1}{N} \sum_i \{g'(x'_i\beta_0)\}^2 x_ix'_i, \quad \text{since } E[y_i - g(x'_i\beta_0)] = 0.
\]

(c) Combining

\[
\sqrt{N}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, A_0^{-1}B_0A_0^{-1})
\]

\[
\xrightarrow{d} \mathcal{N}(0, \left( \lim \frac{1}{N} \sum_i V[y_i|x_i]\{g'(x'_i\beta_0)\}^2 x_ix'_i \right)^{-1} \left( \lim \frac{1}{N} \sum_i \{g'(x'_i\beta_0)\}^2 x_ix'_i \right)^{-1}].
\]

(d) For the example in question 1: \( g(x'_i\beta_0) = \exp(x'_i\beta_0) \), so \( g'(x'_i\beta_0) = \exp(x'_i\beta_0) \).

Also \( V[y_i|x_i] = \exp(x'_i\beta_0)(1 + 2\exp(x'_i\beta_0)) \). So

\[
\sqrt{N}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, A_0^{-1}B_0A_0^{-1})
\]

\[
\xrightarrow{d} \mathcal{N}(0, \left( \lim \frac{1}{N} \sum_i e^{2x'_i\beta_0} x_ix'_i \right)^{-1} \left( \lim \frac{1}{N} \sum_i e^{2x'_i\beta_0}(1 + 2e^{x'_i\beta_0})x_ix'_i \right) (same)^{-1}].
\]

(e) Reject \( H_0: \beta_2 = 0 \) against \( H_a: \beta_2 \neq 0 \) at level \( a \) if \(|t| > 1.96\)

where \( t = \frac{\hat{\beta}_j}{s_{\beta_j}} \) where \( s_{\beta_j}^2 \) is the \( j^{th} \) diagonal entry in \( \left( \sum_i e^{3x'_i\hat{\beta} x_ix'_i} \right)^{-1} \left( \sum_i e^{2x'_i\hat{\beta} x_ix'_i} \right) \left( \sum_i e^{3x'_i\hat{\beta} x_ix'_i} \right)^{-1} \).
3. (a) Simplest is Khinchine's theorem. Let \( \{X_i\} \) be i.i.d. with \( \text{E}[X_i] = \mu \). Then \( \bar{X}_n - \mu \xrightarrow{p} 0. \) [Other LLN's can be given].

(b) Here

\[
Q(\beta) = \sum_i \{y_i x'_i \beta - \exp(x'_i \beta) - \ln y_i!\}
\]

\[
\frac{\partial Q(\beta)}{\partial \beta} = \sum_i \{y_i x_i - \exp(x'_i \beta) x_i\} = \sum_i (y_i - \exp(x'_i \beta)) x_i
\]

\[
\frac{\partial^2 Q(\beta)}{\partial \beta \partial \beta'} = -\sum_i \exp(x'_i \beta) x_i x'_i
\]

Using Newton-Raphson the update rule is

\[
(\hat{\beta}_{s+1} - \hat{\beta}_s) = -\left[ -\sum_i \exp(x'_i \hat{\beta}_s) x_i x'_i |\hat{\beta}_s \right]^{-1} -\sum_i (y_i - \exp(x'_i \hat{\beta}_s)) x_i |\hat{\beta}_s
\]

\[
= \left[ \sum_i \exp(x'_i \hat{\beta}_s) x_i x'_i \right]^{-1} \sum_i (y_i - \exp(x'_i \hat{\beta}_s)) x_i
\]

4. Note: Question 4 was out of 4 points (typo on exam).

(1) We expect the slope coefficient to be consistently estimated, because for this d.g.p. \( \text{E}[y|x] = \exp(\beta_1 + \beta_2 x) \) and NLS estimates with this functional form for the conditional mean. So we expect the average of the slope coefficients to approximately equal \( \beta_2 = 1 \). This is the case here with average of \( b_{2f} \) equal to 1.0033.

(2) We expect the standard errors to be inconsistently estimated however. The default NLS assumes homoskedasticity, but here the d.g.p. is Poisson with \( \text{V}[y|x] = \text{V}[y|x] = \exp(\beta_1 + \beta_2 x) \) which is heteroskedastic.

Thus the standard errors are wrong, with average of \( \text{se}_{2f} \) equal to 0.0413 compared to standard deviation of \( b_{2f} \) equal to 0.075.

(3) This underestimation of the standard errors leads to too high t-statistics and over-rejection of \( H_0: \beta_2 = 1 \). The average of \( \text{reject}_{2f} \) equals 0.272, much more than 0.05.

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.

<table>
<thead>
<tr>
<th>Scores out of</th>
<th>35</th>
<th>A</th>
<th>31 and above</th>
</tr>
</thead>
<tbody>
<tr>
<td>75th percentile</td>
<td>31.5 (90%)</td>
<td>A-</td>
<td>26 and above</td>
</tr>
<tr>
<td>Median</td>
<td>30 (86%)</td>
<td>B+</td>
<td>21 and above</td>
</tr>
<tr>
<td>25th percentile</td>
<td>26 (74%)</td>
<td>B</td>
<td>16 and above</td>
</tr>
</tbody>
</table>