LNL Assignment 1

Meeny Ravi

January 23 2021

1 Question 1

1.1 Part a

$$f(y) = \sqrt{\frac{\lambda}{2\pi y^3}} exp(\frac{-\lambda(y-\mu)^2)}{2y\mu^2} = exp(\ln(f(y))) \tag{1}$$

First, simplifying the squared portion of the formula above:

$$\frac{-\lambda(y-\mu)^2}{2y\mu^2} \tag{2}$$

Expand the exponential term inside the parentheses

$$= \frac{-\lambda(y^2 - 2y\mu + \mu^2)}{2y\mu^2} \tag{3}$$

Distribute the lambda

$$= \frac{-\lambda y^2 + 2\lambda y\mu - \lambda\mu^2}{2y\mu^2} \tag{4}$$

Since the second and third terms of the numerator share common terms, we can combine and simplify

$$= \frac{2\lambda y\mu - \lambda\mu^2}{2y\mu^2} - \frac{\lambda y^2}{2y\mu^2} \tag{5}$$

We can simplify the first term of eq 5 by pulling out a 2 and mu:

$$= \frac{2\mu(\lambda y - \frac{1}{2}\lambda\mu)}{2\mu(y\mu)} - \frac{\lambda y^2}{2y\mu^2} \tag{6}$$

The 2*mu can cancel out and the y in the second term can also cancel

$$= \frac{(\lambda y - \frac{1}{2}\lambda\mu)}{(y\mu)} - \frac{\lambda y}{2\mu^2} \tag{7}$$

Now, in the original, when we logged both side, the other prt of the original equation becomes:

$$\frac{\log(\lambda)}{2} - \frac{\log(2\pi y^3)}{2} \tag{8}$$

Now, combined eq 7 and eq 8:

$$= \frac{(\lambda y - \frac{1}{2}\lambda\mu)}{(y\mu)} - \frac{\lambda y}{2\mu^2} + \frac{\log(\lambda)}{2} - \frac{\log(2\pi y^3)}{2} \tag{9}$$

The first term can be split up like this:

$$= \frac{\frac{-1}{2}\lambda}{y} + \frac{\lambda}{\mu} - \frac{\lambda y}{2\mu^2} + \frac{\log(\lambda)}{2} - \frac{\log(2\pi y^3)}{2} \tag{10}$$

$$= \frac{-\lambda}{2y} + \frac{\lambda}{\mu} - \frac{\lambda y}{2\mu^2} + \frac{\log(\lambda)}{2} - \frac{\log(2\pi y^3)}{2} \tag{11}$$

After moving mu to the numerator and lambda to the denominator, and simplifying we get:

$$= \frac{y(\frac{-1}{\mu^2}) + \frac{2}{\mu}}{\frac{2}{\lambda}} + \frac{-\lambda}{2y} + \frac{\log(\lambda)}{2} - \frac{\log(2\pi y^3)}{2} \tag{12}$$

$$\boxed{= \frac{y(\frac{-1}{\mu^2}) + \frac{2}{\mu}}{\frac{2}{\lambda}} - \frac{\lambda}{2y} + \frac{\log(\lambda)}{2} - \frac{\log(2\pi y^3)}{2}} \tag{13}$$

To compute the values below, I just matched it to the format of the probability density function given

Theta: The value right after y in eq 13

$$\theta = \frac{-1}{\mu^2} \tag{14}$$

Phi: The denominator of the first term if eq 13

$$\phi = \frac{2}{\lambda} \tag{15}$$

a(phi): phi

$$a(\phi) = \phi = \frac{2}{\lambda} \tag{16}$$

b(theta): -2/mu which can be simplified in terms of theta because we have theta in terms of mu in eq 14:

$$\begin{aligned}
b(\theta) &= \frac{-2}{\mu} \\
\theta &= \frac{-1}{\mu^2} \\
&= \mu^2\theta = -1 \\
&= \mu^2 = \frac{1}{-\theta} \\
&= \mu = \frac{1}{\sqrt{-\theta}} \\
b(\theta) &= \frac{-2}{\frac{1}{\sqrt{-\theta}}} \\
b(\theta) &= -2\sqrt{-\theta}
\end{aligned} \tag{17}$$

c(y,phi): The remaining portion of eq 13. We can replace lambda in terms of phi because we have phi in terms of lambda

$$\begin{aligned}
\phi &= \frac{2}{\lambda} \\
\lambda\phi &= 2 \\
\lambda &= \frac{2}{\phi} \\
&\frac{-\lambda}{2y} \\
&= \frac{-\frac{2}{\phi}}{2y} \\
&= \frac{-1}{\phi y} \\
&= \frac{-1}{\phi y} + \frac{\log(\frac{2}{\phi})}{2} - \frac{\log(2\pi y^3)}{2} \\
c(y,\phi) &= \frac{-1}{\phi y} + \frac{\log(\frac{2}{\phi}) - \log(2\pi y^3)}{2}
\end{aligned} \tag{18}$$

$$\boxed{\begin{aligned}
\theta &= \frac{-1}{\mu^2} \\
\phi &= \frac{2}{\lambda} \\
a(\phi) = \phi &= \frac{2}{\lambda} \\
b(\theta) &= -2\sqrt{-\theta} \\
c(y,\phi) &= \frac{-1}{\phi y} + \frac{\log(\frac{2}{\phi}) - \log(2\pi y^3)}{2}
\end{aligned}} \tag{19}$$

1.2 Part b

To get the canonical link:

$$g(\mu_i) \tag{1}$$

$$\mu_i = b'(\theta_i) \tag{2}$$

$$b(\theta) = -2\sqrt{-\theta} \tag{3}$$

Find the derivative and put theta in terms of mu (b')

$$\begin{aligned}
b'(\theta) &= \frac{1}{\sqrt{-\theta}} \\
\mu &= \frac{1}{\sqrt{-\theta}} \\
\mu\sqrt{-\theta} &= 1 \\
\sqrt{-\theta} &= \frac{1}{\mu^2} \\
\theta &= -\frac{1}{\mu^2} \\
&\boxed{g(\mu_i) = -\frac{1}{\mu_i^2}}
\end{aligned} \tag{4}$$

1.3 Part c

In this exploration of risk information search via a search engine: Queries and clicks in healthcare and information security project, it was found that the general public is using search engines to seek information on risks and threats. This paper investigates session length of search and query click rate and found that session length can be characterized well by the Inverse Gaussian distribution.

It is suggested that users' search pattern is affected by search topics, users' activity level, and their interaction. In order to test these categories, the mean parameter of an Inverse Gaussian distribution as the function of search topics, users' search engine activity level, and their interaction was modeled.

Based on a search log from a three month long search engine data, the validity of inverse Gaussian distribution to predict the number of queries and clicks in a search session and show that it provides a strong fit to the data was verified.

Reference Link: `https://www.sciencedirect.com/science/article/pii/S016792361100162X?casa_token=2YZsfbF5rLwAAAAA:dfwHzO4O1Gq7RJglTsksHxSDJMJ9_HMUtF_Ca8565VybwG_wSI4Gc8f2nH_EZaK5UguAFc4rQ_w`

2 Question 2

2.1 Part a

Given log likelihood equation:

$$L(\beta, y, X) = \sum_{i=1}^{n}(y_i(x_i^t\beta) - exp(x_i^t\beta) - \ln(y_i!)) \tag{1}$$

Take first derivative of log likelihood equation with respect to B

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{n}(y_i - exp(x_i^t\beta))x_i \tag{2}$$

set x to 1

$$= \sum_{i=1}^{n}(y_i - exp(\beta)) \tag{3}$$

Set this derivative equal to 0 and solve for B

$$\sum_{i=1}^{n}(y_i - exp(\beta)) = 0 \tag{4}$$

$$\sum_{i=1}^{n}(y_i) = \sum_{i=1}^{n} exp(\beta) \tag{5}$$

Common sum formula where the part in the summation is equal to n*inside summation

$$\sum_{i=1}^{n}(y_i) = n * exp(\beta) \tag{6}$$

Divide by n

$$\frac{1}{n} * \sum_{i=1}^{n}(y_i) = exp(\beta) \tag{7}$$

ln both sides

$$\ln(\frac{1}{n} * \sum_{i=1}^{n}(y_i)) = \beta \tag{8}$$

This is ln of the mean of the range of y values from 1 to n. We un-long this to get the value of beta that maximizes the likelihood

$$\boxed{\beta = \frac{1}{n} * \sum_{i=1}^{n}(y_i)} \tag{9}$$

This is the mean of the range of y values from i=1 to n.It is the average number of occurrences in this interval.

2.2 Part b

I set eq 9 from part a as mu, the average of the occurrences

$$\mu = \frac{1}{n} * \sum_{i=1}^{n}(y_i) \tag{1}$$

Now I plug mu into the log likelihood equation

$$\begin{gathered} L(\beta, y, X) = \sum_{i=1}^{n}(y_i(\mu) - exp(\mu) - \ln(y_i!)) \\ \boxed{L(\beta, y, X) = \sum_{i=1}^{n}(y_i(\mu) - exp(\mu) - \ln(y_i!)} \end{gathered} \tag{2}$$

3 Question 3

3.1 Part a

I retained 6 complete cases. When I ran the algorithm and produced the iteration table below, I see that the iteration converged within 4 iterations.

3.2 part b

Attached figure 1

3.3 part c

Parameter Estimates

Parameter Estimates	
Intercepts	0.128954
AGE	-0.0080593
CAR_AGE	-0.0255993
HOMEKIDS	0.0576943
KIDSDRIV	0.1741679
MVR_PTS	0.1264452
TIF	-0.0299979
TRAV_TIME	0.0052007
YOJ	-0.0161642

Asymptotic Standard Errors

Column	Standard Error
Intercept	0.080473
AGE	0.0015469
CAR_AGE	0.0021769
HOMEKIDS	0.0125402
KIDSDRIV	0.0215512
MVR_PTS	0.0044827
TIF	0.0030273
TRAVTIME	0.0007229
YOJ	0.0026727

Confidence intervals

Confidence Intervals		
columns	lowerlimit	upperlimit
Intercept	-0.028773	0.2866811
AGE	-0.0110913	-0.0050274
CAR_AGE	-0.0298661	-0.0213326
HOMEKIDS	0.0331156	0.082273
KIDSDRIV	0.1319276	0.2164081
MVR_PTS	0.117659	0.1352313
TIF	-0.0359315	-0.0240644
TRAVTIME	0.0037837	0.0066176
YOJ	-0.0214028	-0.0109256

3.4 part d

Attached below figure 2

3.5 part e

From the asymptotic correlation matrix, I see that Age and car age are negatively correlated. Older people have newer cars. The age of the home kids is positively correlated while the age of the driving kids is negatively correlated. Older kids are more at home and the younger kids are mostly driving kids. More kids at home means less kids driving because home kids and driving kids are negatively correlated. The travel time of the homekids is positively correlated while the travel time of the driving kids is negatively correlated. So, the homekids take longer to travel and the driving kids takes less time to travel.

3.2 part b figure 1									
Iteration History									
ITERATION	LOG LIKELIHOOD	Intercept	AGE	CAR_AGE	HOMEKIDS	KIDSDRIV	MVR_PTS	TIF	TRAVTIME	YOJ
0	-14055.13228	-0.2166234	0	0	0	0	0	0	0	0
1	-13389.58608	0.1409492	-0.0074564	-0.0245483	0.0687899	0.231537	0.1585273	-0.0277757	0.0050182	-0.0190219
2	-13287.43075	0.1224621	-0.007952	-0.0254869	0.0591542	0.1847715	0.1314797	-0.0297325	0.0051719	-0.0165777
3	-13285.98557	0.1287989	-0.0080571	-0.0255984	0.0577127	0.1744252	0.1265368	-0.0299952	0.0052002	-0.0161707
4	-13285.98509	0.128954	-0.0080593	-0.0255993	0.0576943	0.174168	0.1264452	-0.0299979	0.0052007	-0.0161642
5	-13285.98509	0.128954	-0.0080593	-0.0255993	0.0576943	0.1741679	0.1264452	-0.0299979	0.0052007	-0.0161642
6	-13285.98509	0.128954	-0.0080593	-0.0255993	0.0576943	0.1741679	0.1264452	-0.0299979	0.0052007	-0.0161642

3.4 part d figure 2									
Asymptotic Correlation Matrix									
column	Intercept	AGE	CAR_AGE	HOMEKIDS	KIDSDRIV	MVR_PTS	TIF	TRAVTIME	YOJ
Intercept	1	-0.8190333	-0.114287	-0.480248	0.1957051	-0.2206166	-0.189584	-0.3145274	-0.1605653
AGE	-0.8190333	1	-0.1017206	0.4699115	-0.2243712	0.070636	0.0178565	-0.006691	-0.1755531
CAR_AGE	-0.114287	-0.1017206	1	0.0942214	-0.0005582	0.0325112	-0.0047667	0.0215871	-0.0583403
HOMEKIDS	-0.480248	0.4699115	0.0942214	1	-0.515173	-0.0321312	-0.0000477	0.0207216	-0.1276512
KIDSDRIV	0.1957051	-0.2243712	-0.0005582	-0.515173	1	-0.0251396	0.0235815	-0.0274045	0.0175536
MVR_PTS	-0.2206166	0.070636	0.0325112	-0.0321312	-0.0251396	1	0.0269478	0.0025015	0.047987
TIF	-0.189584	0.0178565	-0.0047667	-0.0000477	0.0235815	0.0269478	1	0.0006074	-0.0323721
TRAVTIME	-0.3145274	-0.006691	0.0215871	0.0207216	-0.0274045	0.0025015	0.0006074	1	0.0014582
YOJ	-0.1605653	-0.1755531	-0.0583403	-0.1276512	0.0175536	0.047987	-0.0323721	0.0014582	1