Question 4

a) \[ p(y|d,v) = \int \frac{e^{-\lambda} \mu^{y-1} e^{-\lambda}}{\Gamma(y) \lambda^y} \, d\lambda \]

\[ = \frac{1}{\Gamma(y) \lambda^y} \int \mu^{y-1} e^{-\lambda (1+\frac{1}{2})} \, d\lambda \]

\[ = \frac{\Gamma(y+v)}{\Gamma(y) \lambda^y (1+\frac{1}{2})^{y+v}} \]

\[ = \frac{\Gamma(y+v)}{\Gamma(y) \lambda^y} \left( \frac{1}{1+\lambda} \right)^y \left( \frac{1}{1+\lambda} \right)^v \]

which we recognise as being Negative Binomial \((y, \frac{1}{1+\lambda})\)

hence \(EY = \frac{\nu(1+\lambda)}{(1+\lambda)} = \nu \lambda\).

\(\text{Var}(Y) = \frac{\nu(1+\lambda)}{(1+\lambda)^2} = \nu + \frac{\nu}{\lambda^2}\).

Now consider \(EY_i = \frac{\mu_i}{\nu} \) but \(Y_i = Y + \epsilon_i\), we re-parameterise \(\epsilon_i = \frac{\mu_i}{\nu}\) hence \(\text{Var}(Y_i) = \frac{\nu}{\lambda^2} + \frac{\nu}{\lambda^2} \nu_i^2\).

We also observe that for \(\nu\) fixed we can write

\[ p(y_i|d,Y) = \exp \left\{ \frac{\mu_i}{\nu} \log \left( \frac{\mu_i}{\nu} \right) + \nu \log \left( \frac{1}{1+\frac{\mu_i}{\nu}} \right) - \log \left( \frac{\Gamma(y_i+\nu)}{\Gamma(y_i) \lambda^y} \right) \right\} \]

which is in exponential family form with

\(\Theta_i = \log \left( \frac{\mu_i}{\nu} \right), \quad b(\Theta_i) = -\nu \log \left( 1 - e^{\Theta_i} \right)\)

\(\mathbf{a}(\phi) = 1 \quad \text{and} \quad \mathbf{a}_{\phi} = -\log \left( \frac{\Gamma(y_i+\nu)}{\Gamma(y_i) \lambda^y} \right)\).
Conversely, setting $d_i = d$, constant,

$$ P(y; d, Y_i) = \exp \{ \log \left( \frac{Y_i \exp(y)}{Y_i} \right) + y \log \left( \frac{1}{d} \right) + y \log \left( \frac{d}{Y_i} \right) - \log Y_i \} $$

which cannot be written in exponential family form due to the term $Y_i(Y_i + d)$.

In this case, re-parametrizing $X_i = \frac{y_i}{d}$, we have $\text{var}(Y_i) = (1+d)X_i$, $i.e.$ an over-dispersed Poisson variance.

Finally, if $d_i = \theta + \psi X_i$ and $V_i = \theta + \psi X_i$, then:

$$ E_y = \frac{\theta + \psi X_i}{1 + \theta X_i} = X_i, \quad \text{Var}(y) = \mu_i + \mu_i(\theta + \psi X_i) $$

$$ = \mu_i + \theta \mu_i + \psi \mu_i^2 $$

Note that since the $V_i$ change with $X_i$, this cannot be written in exponential family form for the same reasons as above.

Finally, for $\theta = 1, \theta = 1$, the quasi-likelihood is

$$ Q(\theta, y) = \int_y^\infty \frac{y \log \left( \frac{1}{\theta X_i} \right)}{y \log \left( \frac{1}{\theta X_i} \right)} \, dx = \int_y^\infty \frac{1}{1 + \theta X_i} \, dx = y \log \left( \frac{\theta X_i}{y + \theta} \right) - \frac{\theta X_i}{y + \theta} $$

which is negative binomial, up to a constant. For $V(\mu) = (1+d)\mu$,

$$ Q(\theta, y) = \int_y^\infty \frac{1}{1 + \theta X_i} \, dx = \frac{y \log \left( \frac{1}{\theta} \right) - \frac{\theta X_i}{1 + \theta}}{1 + \theta} $$

$$ = \frac{1}{1 + \theta} \left( y \log \mu - \mu \right) $$

which is a scaled version of the Poisson likelihood.
Any of the variance parameters could be fit in one of two ways:

1. Since we always have a negative binomial model, we can maximize a likelihood. However, this cannot always be done within a 
   framework.

2. In a quasi-likelihood setting, we could consider solving

   \[ \frac{\sum_{i=1}^{n} (y_i - \mu_i)^2}{n \sigma^2} - 1 = 0 \]

   where we have subsumed variance parameters into \( \sigma^2 \).

The models could be distinguished by comparing their likelihoods if the first approach in part 5 was taken.

Alternatively, forming the squared raw residuals \( z_i = (y_i - \mu_i)^2 \), we could examine the significance of linear and quadratic terms in the regression \( (z_i - \mu_i) = \beta_0 + \beta_1 \mu_i + \beta_2 \mu_i^2 \), possibly appropriately weighted.
Question 3

a) We observe that if \( A \) and \( B \) are symmetric and positive definite then \( A^{-1}B \) and \( B^{-1}A \) are also positive definite.

Then, if \( A - B \succeq 0 \), \( A^{-1}(A-B)B^{-1} \succeq 0 \Rightarrow B^{-1}A^{-1} \succeq 0 \).

b) If \( \beta^* \) is a solution to \( D^T V^{-1}(y - mu) = 0 \)
then \( \text{cov}(\beta^*) = (D^T V^{-1}D)^{-1} = A \)
and \( \beta^* \) is the covariance of \( \beta \) and \( V = V_0 \) this reduces to \( \text{cov}(\beta^*) = (D^T V^{-1}D)^{-1} = A \).

Now, we consider
\[
B^{-1}A^{-1} = D^T (V_0^{-1} - V_0^{-1}D(D^T V^{-1}V_0^{-1}D)D^T V^{-1}) D
\]
and we observe that this is the covariance of
\[
(I - V_0^{-1}D(D^T V^{-1}V_0^{-1}D)D^T V^{-1}) y \quad \text{and} \quad D^T V^{-1}y.
\]
and is positive definite. Hence \( A - B \succeq 0 \).

Question 4

If we write \( V = D R D \) with \( D \) diagonal, \( R \) independent of \( y \) and assume \( V \) is positive definite, then \( D \) and \( R \) are also positive definite and hence the diagonal entries of \( D \) are strictly positive.

Now, we observe that from \( \mu \text{ln} pg.334 \), with
\( L = V^{-1} \), \( W_{ij} = \frac{r_{i}^{*} y_{j}}{\text{det}(D) \text{det}(W)} \) and \( r_{ij}^{*} = (y_{i}, y_{j}) \text{entry of } R^{-1} \)
the existence of a quasi-likelihood requires
\[
\frac{\partial W_{ij}}{\partial \mu_{i}} = \frac{\partial W_{ij}}{\partial \mu_{j}} = \frac{\partial W_{ij}}{\partial \mu_{k}} \quad \text{for } i, j, k.
\]
Now, \( \frac{\partial W_{ii}}{\partial \mu_{j}} = 0 = \frac{\partial W_{ij}}{\partial \mu_{j}} = \frac{\partial W_{ij}}{\partial \mu_{j}} \frac{d \text{det}(W)}{d \mu_{j}} \frac{1}{d y_{i}} \frac{1}{d y_{j}} \frac{r_{ij}^{*}}{d y_{i}^{2}} \frac{d y_{i}}{d y_{j}} \)

since \( d y_{i}^{2} > 0 \) this is only true if \( r_{ij}^{*} = 0 \) (i.e. \( R \) is diagonal) or \( d \text{det}(W) / d \mu_{j} = 0 \) i.e. \( D \) does not depend on \( \mu \).