Problem 1

A study is conducted to quantify the evidence against the null hypothesis that less than 80 percent of the Swiss population have antibodies against the human herpesvirus. Among a total of 117 persons investigated, 105 had antibodies.

1. In this problem the goal is to quantify the proportion of people in the population that have antibodies. One wants to test the hypothesis

$$H_0 : p < 0.8 \text{ vs. } H_1 : p \geq 0.8,$$

where p is the population proportion. A reasonable statistic to use, given the data, is $\hat{p} = 105/117 \approx 0.897$. The question is if it is possible to derive an optimal test for this hypothesis. An appropriate distribution for the question is the binomial distribution. Define the random variable $X \sim Bin(n, p)$, where $n = 117$ are the number of trials, p is the probability of success (a person has antibodies) in each trial and there are $x = 105$ successes in the sample. It is known that the binomial distribution belongs to the exponential family of distributions, which means that it has a monotone likelihood ratio. The Karlin-Rubin theorem thus implies that there exists an optimal test for this hypothesis for any significance level α. Specifically, the test that rejects H_0 iff $T > t_0$, for any t_0, is a UMP level α test where $\alpha = \mathrm{P}_p(T > t_0)$ and T is a sufficient statistic for p.

2. The exact null distribution from the binomial model is

$$f(x; p) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} (1-p)^n e^{x \log\left(\frac{p}{1-p}\right)}.$$

Thus, it is apparent that the sufficient statistic in this case is $T(x) = x$. The p-value is defined as

$$\mathrm{p}(x) = \sup_{p<0.8} \mathrm{P}_p(T(X) \geq T(x)),$$

where x is the observed value in the sample. This means that the p-value is

$$\mathrm{p}(115) = \mathrm{P}(X \geq 105 | p = 0.8) \approx 0.00155,$$

using the function **pbinom** in R to calculate the CDF of the binomial.

For instance, setting the significance level to $\alpha = 0.05$ would lead us to conclude that H_0 should be rejected.

The advantage of this procedure is that the result is exact, i.e. since we use the exact distribution and do not apply (for example) an asymptotic approximation, this result is trustworthy. The disadvantage of this procedure is that the exact null distribution has to be known, because the method cannot be used if it is unknown. In such cases, one could use asymptotic approximations, which are used in the following. Another advantage of this method is that it does not require a lot of computation compared to some of the test statistics that are based on asymptotic properties of the maximum likelihood estimator.

3. The Wald statistic for this problem is defined as either of the two following statistics

$$W_1 := \sqrt{I(\hat{p}_{ML})}(\hat{p}_{ML} - p^*), \tag{1}$$
$$W_2 := \sqrt{\mathcal{I}(\hat{p}_{ML})}(\hat{p}_{ML} - p^*), \tag{2}$$

where I is the observed Fisher information, $\mathcal{I}$ is the expected Fisher information and $\hat{p}_{ML}$ is the maximum likelihood estimator of p. Both of these statistics are asymptotically standard normally distributed when p^* is the true parameter value.

It is already known that the maximum likelihood estimator of p is $\hat{p}_{ML} = \frac{x}{n}$, but this will be shown quickly before proceeding. First of all, the likelihood function is

$$L(p|x) = p^x(1-p)^{n-x},$$

where the binomial coefficient has been ignored, since it does not contain the parameter of interest. The log-likelihood is

$$l(p|x) = x \log p + (n-x)\log(1-p),$$

which means that the score function is

$$\mathcal{S}(p) = \frac{\mathrm{d}}{\mathrm{d}p} l(p|x) = \frac{x}{p} + \frac{x-n}{1-p} = \frac{x-np}{p(1-p)},$$

and the observed Fisher information is

$$I(p) = -\frac{\mathrm{d}^2}{\mathrm{d}p^2} l(p|x) = \frac{x}{p^2} - \frac{x-n}{(1-p)^2}.$$

Setting the score function to zero yields

$$\hat{p}_{ML} = \frac{x}{n},$$

which we already used in part **1.** when defining a reasonable statistic for the hypothesis test. The second derivative at $\hat{p}_{ML}$ is

$$-I(\hat{p}_{ML}) = -\frac{n^3}{x(n-x)} < 0,$$

which shows that $\hat{p}_{ML}$ indeed is the MLE of p. Moreover, note that the expected fisher information is

$$\mathcal{I}_p(p) = -\mathrm{E}(I(p)) = \frac{1}{p^2}\mathrm{E}(x) - \frac{1}{(1-p)^2}(\mathrm{E}(x) - n) = \frac{n}{p} + \frac{n}{1-p} = \frac{n}{p(1-p)}.$$

The first Wald statistic, as shown in equation (1), gives

$$W_1 = \sqrt{I(\hat{p}_{ML})}(\hat{p}_{ML} - p^*) = \sqrt{\frac{n^3}{x(n-x)}}\left(\frac{x}{n} - p^*\right) \sim N(0,1),$$

which means that

$$\begin{aligned} & -z_{1-\alpha/2} \leq \sqrt{\frac{n^3}{x(n-x)}}\left(\frac{x}{n} - p^*\right) \leq z_{1-\alpha/2} \\ \implies & \frac{x}{n} - z_{1-\alpha/2}\left(\frac{n^3}{x(n-x)}\right)^{-1/2} \leq p^* \leq \frac{x}{n} + z_{1-\alpha/2}\left(\frac{n^3}{x(n-x)}\right)^{-1/2} \end{aligned}$$

is a (asymptotic) $(1-\alpha)100\%$ confidence interval for p. Also, the p-value can be calculated as

$$\begin{aligned} \mathrm{p} &= \sup_{p<0.8} \mathrm{P}_p(W_1 \geq w_1) \\ &= \mathrm{P}_p(W_1 \geq w_1 | p = 0.8) \\ &\approx 1 - \Phi(3.473862) \approx 0.000257, \end{aligned}$$

where $\Phi(x)$ is the CDF of the standard normal distribution.

For instance, setting $\alpha = 0.05$ gives a 95% confidence interval for p

$$p^* \in [0.8425, 0.9524],$$

which would make us conclude that H_0 should be rejected, which is the same conclusion as in **2.** when using a significance level of $\alpha = 0.05$. Note that the same conclusion is found from the p-value above.

4. The logit transformation is

$$\tau(p) = \log\left(\frac{p}{1-p}\right).$$

We know that the MLE is invariant to transformations, which means that

$$\hat{\tau}_{ML} = \log\left(\frac{\hat{p}_{ML}}{1-\hat{p}_{ML}}\right) = \log\left(\frac{x}{n-x}\right).$$

The expected Fisher information of $\tau(p)$ is

$$\mathcal{I}_\tau(\tau(p)) = \mathcal{I}_p(p) \left\{ \frac{\mathrm{d}\tau(p)}{\mathrm{d}p} \right\}^{-2} = \frac{n}{p(1-p)} \left\{ \frac{1}{p(1-p)} \right\}^{-2} = np(1-p).$$

Thus, the second Wald statistic for the logit-transformation, following equation (2), is

$$W_2 = \sqrt{\mathcal{I}_\tau(\hat{p}_{ML})}(\hat{\tau}_{ML} - \tau^*),$$

which is asymptotically standard normally distributed when τ^* is the true value of the transformed parameter. This means that the observed Wald statistic is

$$w_2|(p \in H_0) = \sqrt{n\frac{x}{n}(1-\frac{x}{n})} \left(\log\left(\frac{x}{n-x}\right) - \log\left(\frac{p}{1-p}\right)\right) \Bigg| (p \in H_0)\,.$$

Finally, this leads to the p-value

$$\begin{aligned} \mathrm{p} &= \sup_{p<0.8} \mathrm{P}_p(W_2 \geq w_2) \\ &= \mathrm{P}_p(W_2 \geq w_2 | p = 0.8) \\ &\approx 1 - \Phi(2.568743) \approx 0.0051. \end{aligned}$$

For instance, when setting the significance level to $\alpha = 0.05$, this leads to the same conclusion as in the earlier cases.

5. The score statistic is used to obtain a p-value. In the uniparametric case, the score statistic is defined as either of the two following statistics

$$S_1 := \frac{\mathcal{S}(p^*)^2}{I(p^*)}, \tag{3}$$

$$S_2 := \frac{\mathcal{S}(p^*)^2}{\mathcal{I}(p^*)}. \tag{4}$$

When p^* is the true parameter value, both these statistics are distributed as χ^2_1. The expression of S_2 is

$$S_2(p) = \frac{\left(\frac{x-np}{p(1-p)}\right)^2}{\frac{n}{p(1-p)}} = \frac{(x-np)^2}{np(1-p)}$$

This means that the p-value is

$$
\begin{aligned}
\mathrm{p} &= \sup_{p<0.8} \mathrm{P}_p(S_2 \geq s_2) \\
&= \mathrm{P}_p(W_2 \geq s_2 | p = 0.8) \\
&\approx 1 - C(6.942308) \approx 0.0084,
\end{aligned}
$$

where C is the CDF of χ^2_1.

For instance, when setting the significance level to $\alpha = 0.05$, this leads to the same conclusion as in the earlier cases.

We don't need to consider transformations in this case because the score statistic is invariant to reparameterizations of p. This can be shown relatively easily, with some tedious calculations. Let $\tau(p)$ be a reparameterization of p. It can be shown that $\mathcal{S}(\tau(p)) = \mathcal{S}(p)\left\{\frac{\mathrm{d}\tau(p)}{\mathrm{d}p}\right\}^{-1}$. This, together with the transformation of the expected Fisher information can be used to show that $S_2 = \frac{\mathcal{S}(\tau(p))^2}{\mathcal{I}(\tau(p))} = \frac{\mathcal{S}(p)^2}{\mathcal{I}(p)}$.

6. Some comments on the approximations taken above are made. First of all, the exact null hypothesis does not use an asymptotic approximation, which means that this yields the most trustworthy results in this case. Secondly, all the approximations that are made above are asymptotic, which means that they are good approximations for a large sample size n. Thus, an important question is if $n = 117$ is large enough for these approximations to be good? This is not straightforward to answer. Luckily, the conclusion of the study is the same in all cases when choosing a significance level of for example $\alpha = 0.05$ or $\alpha = 0.01$, which means that we could have used any of these tests in practice in this case. With the two mentioned significance levels there is enough evidence against the null hypothesis in the sample and the conclusion is that 80% or more of the Swiss population have antibodies against human herpesvirus.

Problem 2

In a cohort study on the incidence of ischaemic heart disease (IHD), 337 male probands were enrolled. Each man was categorised as non-exposed (group 1, daily energy consumption ≥ 2750 kcal) or exposed (group 2, daily energy consumption < 2750 kcal) to summarise his average level of physical activity. For each group, the number of person years ($Y_1 = 2768.9$ and $Y_2 = 1857.5$), and the number of IHD cases ($D_1 = 17$ and $D_2 = 28$) was registered thereafter. We assume that

$$
D_i | \lambda_i \sim Poiss(\lambda_i Y_i),\ i = 1, 2,
$$

where $\lambda_i > 0$ is the group-specific incidence rate.

1. The MLE $\hat{\lambda}_i$ is derived for each group. The distribution function for the men in group i is

$$f(d_i;\lambda_i Y_i) = \frac{(\lambda_i Y_i)^{d_i} e^{-\lambda_i Y_i}}{d_i!}.$$

Thus, the likelihood functions are

$$L(\lambda_i; Y_i, d_i) = (\lambda_i Y_i)^{d_i} e^{-\lambda_i Y_i},\ i = 1,2,$$

when we ignore factors that do not depend on the parameter. This means that the log-likelihood functions are

$$l(\lambda_i; Y_i, d_i) = d_i \log(\lambda_i Y_i) - \lambda_i Y_i,\ i = 1,2.$$

The score functions are

$$\mathcal{S}(\lambda_i; Y_i, d_i) = \frac{\mathrm{d}}{\mathrm{d}\lambda_i} l(\lambda_i; Y_i, d_i) = \frac{d_i}{\lambda_i} - Y_i,\ i = 1,2.$$

The roots of the score functions are $\hat{\lambda}_i = \frac{d_i}{Y_i}$. The observed Fisher information is

$$I(\hat{\lambda}_i) = -\frac{\mathrm{d}^2}{\mathrm{d}\lambda_i^2} l(\lambda_i; Y_i, d_i)\Big|_{\hat{\lambda}_i} = \frac{d_i}{\lambda_i^2}\Big|_{\hat{\lambda}_i} = \frac{Y_i^2}{d_i} > 0,\ d_i \neq 0,$$

which shows that the MLE is $\hat{\lambda}_i = \frac{d_i}{Y_i}$, $i = 1,2$. The numerical values are $\hat{\lambda}_1 = \frac{17}{2768.9} \approx 0.00614$ and $\hat{\lambda}_2 = \frac{28}{1857.5} \approx 0.0151$. Note also that the expected Fisher information is

$$\mathcal{I}_\lambda(\lambda_i) = \mathrm{E}\left(\frac{d_i}{\lambda_i^2}\right) = \frac{Y_i}{\lambda_i}.$$

A 95% Wald confidence interval for $\tau(\lambda) = \log(\lambda_i)$ is given next. As noted in **Problem 1**, the Wald statistic is

$$W_2 := \sqrt{\mathcal{I}_\tau(\hat{\lambda}_i)}(\tau(\hat{\lambda}_i) - \tau(\lambda_i^*)),$$

where λ_i^* is the true value of the parameter. Moreover, the expected Fisher information of $\tau(\lambda_i)$ is

$$\mathcal{I}_\tau(\tau(\lambda_i)) = \mathcal{I}_\lambda(\lambda_i)\left\{\frac{\mathrm{d}\tau(\lambda_i)}{\mathrm{d}\lambda}\right\}^{-2} = \frac{Y_i}{\lambda_i}\left\{\frac{1}{\lambda_i}\right\}^{-2} = Y_i\lambda_i.$$

Recall that W_2 is asymptotically standard normally distributed when λ_i^* is the true value of the parameter. This means that we can build a $(1-\alpha)100\%$ confidence interval by

$$
\begin{aligned}
& - z_{1-\alpha/2} \leq \sqrt{\mathcal{I}_\tau(\hat{\lambda}_i)}(\tau(\hat{\lambda}_i) - \tau(\lambda_i^*)) \leq z_{1-\alpha/2} \\
\implies & \tau(\hat{\lambda}_i) - (\mathcal{I}_\tau(\hat{\lambda}_i))^{-1/2} z_{1-\alpha/2} \leq \tau(\lambda_i^*) \leq \tau(\hat{\lambda}_i) + (\mathcal{I}_\tau(\hat{\lambda}_i))^{-1/2} z_{1-\alpha/2}.
\end{aligned}
$$

In our case we are looking for a 95% confidence interval, which gives $z_{1-\alpha/2} = z_{0.975} \approx 1.96$. Also, we have the values $\mathcal{I}_\tau(\hat{\lambda}_1) = 17$ and $\mathcal{I}_\tau(\hat{\lambda}_2) = 28$, which means that $(\mathcal{I}_\tau(\hat{\lambda}_1))^{-1/2} = 0.2425356$ and $(\mathcal{I}_\tau(\hat{\lambda}_2))^{-1/2} = 0.1889822$. Finally, the 95% Wald confidence intervals for $\log(\lambda_i)$ are

$$\log(\lambda_1^*) \in [-5.568, -4.618],$$

and

$$\log(\lambda_2^*) \in [-4.565, -3.824].$$

Back-transformation now yields the confidence intervals

$$\lambda_1^* \in [0.00382, 0.009876],$$

and

$$\lambda_2^* \in [0.0104, 0.0218].$$

2. In order to analyse whether $\lambda_1 = \lambda_2$, the model is reparameterized with $\lambda = \lambda_1$ and $\theta = \frac{\lambda_2}{\lambda_1}$. The joint likelihood is

$$L(\lambda_1, \lambda_2; Y_1, Y_2, d_1, d_2) = (\lambda_1 Y_1)^{d_1} (\lambda_2 Y_2)^{d_2} e^{-\lambda_1 Y_1} e^{-\lambda_2 Y_2},$$

where we have ignored the denominator in the distribution function of the Poisson, since it clearly does not depend on the parameters. This gives the log-likelihood function

$$l(\lambda_1, \lambda_2; Y_1, Y_2, d_1, d_2) = d_1(\log(\lambda_1) + \log(Y_1)) + d_2(\log(\lambda_2) + \log(Y_2)) - \lambda_1 Y_1 - \lambda_2 Y_2.$$

Ignoring the terms that do not depend on the parameters, the kernel of the log-likelihood function is

$$l(\lambda_1, \lambda_2; Y_1, Y_2, d_1, d_2) = d_1 \log(\lambda_1) + d_2 \log(\lambda_2)) - \lambda_1 Y_1 - \lambda_2 Y_2.$$

Finally, reparameterization gives

$$
\begin{aligned}
l(\lambda,\theta) &= d_1\log(\lambda_1) + d_2\log(\lambda_2) - \lambda_1 Y_1 - \lambda_2 Y_2 \\
&= d_1\log(\lambda) + d_2\log(\theta\lambda) - \lambda Y_1 - \theta\lambda Y_2 \\
&= (d_1+d_2)\log(\lambda) + d_2\log(\theta) - \lambda Y_1 - \theta\lambda Y_2 \\
&= d\log(\lambda) + d_2\log(\theta) - \lambda Y_1 - \theta\lambda Y_2,
\end{aligned}
$$

where $d = d_1 + d_2$.

3. The MLE $(\hat{\lambda},\hat{\theta})$, the observed Fisher information matrix $I(\hat{\lambda},\hat{\theta})$ and both profile log-likelihood functions $l_p(\lambda) = l(\lambda,\hat{\theta}(\lambda))$ and $l_p(\theta) = l(\hat{\lambda}(\theta),\theta)$ are calculated here.

   First of all, the MLE can be found via standard calculation, as done in the following. The score vector $\mathcal{S}(\theta,\lambda)$ is

$$
\mathcal{S}(\theta,\lambda)^T = \left(\frac{d}{\lambda} - Y_1 - \theta Y_2, \frac{d_2}{\theta} - \lambda Y_2\right)^T.
$$

   The roots of the score vector are found by equating to zero and solving the system of equations. The solution is $(\hat{\lambda},\hat{\theta}) = (\frac{d-d_2}{Y_1}, \frac{d_2}{d-d_2}\frac{Y_1}{Y_2}) = (\frac{d_1}{Y_1}, \frac{d_2}{d_1}\frac{Y_1}{Y_2}) = (\hat{\lambda}_1, \hat{\lambda}_2\hat{\lambda}_1^{-1}) \approx (0.00614, 2.455)$. This is the MLE because the observed Fisher information matrix (below) is positive semi-definite, which is equivalent to the Hessian being negative semi-definite. Moreover, this is the only critical point inside the parametric space, since this is the only solution of the first order conditions, and the maximum is not reached at the boundary of the parametric space. Note that the invariance principle of the MLE is seen in both coordinates of $(\hat{\lambda},\hat{\theta})$. The Fisher information matrix is

$$
I(\lambda,\theta) = \begin{pmatrix} \frac{d}{\lambda^2} & Y_2 \\ Y_2 & \frac{d_2}{\theta^2} \end{pmatrix}.
$$

   This means that the observed Fisher information is

$$
I(\hat{\lambda},\hat{\theta}) = \begin{pmatrix} \frac{d}{d_1^2}Y_1^2 & Y_2 \\ Y_2 & d_2\left(\frac{Y_2}{Y_1}\right)^2\left(\frac{d_1}{d_2}\right)^2 \end{pmatrix} = \begin{pmatrix} d\hat{\lambda}_1^{-2} & Y_2 \\ Y_2 & d_2\hat{\lambda}_1^2\hat{\lambda}_2^{-2} \end{pmatrix} \approx \begin{pmatrix} 1193793 & 1857.5 \\ 1857.5 & 4.645 \end{pmatrix}.
$$

   The profile log-likelihoods are calculated next. Using that $\hat{\theta}(\lambda) = \frac{d_2}{Y_2}\frac{1}{\lambda}$, we get

$$
\begin{aligned}
l_p(\lambda) = l(\lambda,\hat{\theta}(\lambda)) &= d\log(\lambda) + d_2[\log(d_2) - \log(Y_2) - \log(\lambda)] - \lambda Y_1 - d_2 \\
&\propto d_1\log(\lambda) - \lambda Y_1,
\end{aligned}
$$

   where we have ignored the terms that do not depend on λ. Using that $\hat{\lambda}(\theta) = \frac{d}{\theta Y_2 + Y_1}$, we get

$$
\begin{aligned}
l_p(\theta) = l(\hat{\lambda}(\theta), \theta) &= d[\log(d) - \log(\theta Y_2 + Y_1)] + d_2 \log(\theta) - \frac{dY_1}{\theta Y_2 + Y_1} - \frac{dY_2\theta}{\theta Y_2 + Y_1} \\
&\propto -d\log(\theta Y_2 + Y_1) + d_2 \log(\theta),
\end{aligned}
$$

where we have ignored the terms that do not depend on θ.

4. Done in separate file.

5. A 95% Wald confidence interval for $\log(\theta)$ is computed based on the profile log-likelihood. Ideally, I would like to calculate the score function and Fisher information based on $\log(\theta)$, but the log of the sum above (the term $-d\log(\theta Y_2 + Y_1)$) ruins it for me.

   The score function of $l_p(\theta)$ is

$$
\mathcal{S}_\theta(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta} l_p(\theta) = -\frac{dY_2}{\theta Y_2 + Y_1} + \frac{d_2}{\theta}
$$

   The Fisher information of $l_p(\theta)$ is

$$
I_\theta = \frac{\mathrm{d}}{\mathrm{d}\theta}\mathcal{S}_\theta(\theta) = -\frac{d_2}{\theta^2} + \frac{dY_2^2}{(\theta Y_2 + Y_1)^2}
$$

   The expected Fisher information is ?

   Transforming the expected Fisher information yields

$$
\mathcal{I}(\theta) = ?
$$

   Thus, the Wald statistic is

$$
W_\theta := \sqrt{\mathcal{I}(\hat{\theta})}(\hat{\theta} - \theta^*).
$$

   Since $\theta = \frac{\lambda_2}{\lambda_1}$, the confidence interval for $\log(\theta)$ can be used for inference about the hypothesis $\lambda_1 = \lambda_2$. If the 95% Wald confidence interval for $\log(\theta)$ does not contain 0, it means that we can conclude that $\lambda_1 \neq \lambda_2$ with 95% confidence, i.e. we reject the null-hypothesis that $\lambda_1 = \lambda_2$. The same logic can be used when discussing the p-value for the null-hypothesis; if it is smaller than 0.05, we would reject the null-hypothesis. The p-value can be calculated via

$$
\mathrm{p} = \sup_{\lambda_1 = \lambda_2} \mathrm{P}(W_\theta \geq w_\theta),
$$

   where w_θ is the observed Wald statistic from the sample.

Problem 3

Let $X_1, \dots, X_n$ be a i.i.d. sample from a one-parameter Weibull distribution $W(\alpha, 1)$, with shape parameter α. The MLE of the parameter α cannot be obtained analytically.

First of all, the pdf of $W(\alpha, 1)$ is

$$f(x) = \alpha x^{\alpha-1} e^{-x^\alpha} I_{(0,\infty)}(x). \tag{5}$$

This means that the likelihood function for the sample is

$$L(\alpha|x) = \prod_{i=1}^{n} \alpha x_i^{\alpha-1} e^{-x_i^\alpha} I_{(0,\infty)}(x_i) = \alpha^n e^{-\sum x_i^\alpha} \prod_{i=1}^{n} x_i^{\alpha-1} I_{(0,\infty)}(x_i).$$

1. The log-likelihood function in this problem is

$$l(\alpha|x) = \log L(\alpha|x) = n \log \alpha - \sum_{i=1}^{n} x_i^\alpha + (\alpha - 1) \sum_{i=1}^{n} \log x_i,$$

for $x_{(1)} > 0$ and $l(\alpha|x) = 0$ elsewhere.

2. The partial derivatives of the log-likelihood functions are

$$l'(\alpha|x) = \frac{\partial}{\partial \alpha} l(\alpha|x) = \frac{n}{\alpha} - \sum_{i=1}^{n} x_i^\alpha \log x_i + \sum_{i=1}^{n} \log x_i = \frac{n}{\alpha} + \sum_{i=1}^{n} \log x_i (1 - x_i^\alpha),$$

and

$$l''(\alpha|x) = \frac{\partial^2}{\partial \alpha^2} l(\alpha|x) = -\frac{n}{\alpha^2} - \sum_{i=1}^{n} x_i^\alpha (\log x_i)^2.$$

3. An explicit expression for the first iterates of the Newton-Raphson algorithm for estimating α using the method-of-moments estimator is given here. First of all, the Newton-Raphson algorithm takes the form

$$\alpha_{n+1} = \alpha_n - \frac{l'(\alpha_n|x)}{l''(\alpha_n|x)}.$$

Method-of-moments estimation gives the equation

$$\mu = \Gamma\left(1 + \frac{1}{\alpha}\right),$$

which cannot be solved analytically. Thus, this cannot be used to find the explicit expression. This was supposed to be used as an initial value for the Newton-Raphson fixed point iteration. Instead, another initial value was chosen.

4. Done in separate file.
5. Done in separate file.