

# Statistics 520: Assignment 5

Sam Olson

## Assignment 5

The objectives of this assignment are to (1) ensure that you have a grasp on using the tools of basic likelihood in a data analysis and (2) help you to continue to develop the precise use of notation in presenting descriptions of analyses. On the course web page is a file in the Data folder called `gammadat.txt`. This file contains two columns of values with a header having labels `group1` and `group2`. Each column should be considered to contain values corresponding to a set of independent and identical gamma random variables. That is, the two columns are values from two groups that we wish to compare using a two-sample model with gamma distributions. Consider the first column to contain values for Group 1 and the second column to contain values for Group 2.

A number of resources are available to you to help you complete this assignment. Chapter 5 of the course notes contains a summary of likelihood methods. In the Computing folder of the course web page is a file `newtraph.txt` that contains a generic Newton-Raphson algorithm that you may use for maximum likelihood estimation. There is also a file called `newtraphexplain.txt` that describes the inputs needed, the syntax, and the output. Alternatively you may choose to make use of the built-in R functions `optim` or `nlm`. Any of these options (or others you might know of if you prefer Matlab or something else) are fine as long as you know what you are doing and can produce the quantities needed to conduct the analysis.

Your answer should contain complete and consistent notation using no undefined symbols. You should always clearly explain what you computed and the formulas used. Your answer should not contain computer code or material from a “screen dump.” You will not be awarded any points for such material. If you want to report estimated values do so in the text, as a list, or construct a table.

Again, do not include copied computer function output. You will not get credit for anything presented in that way.

1.

Assume random variables  $Y_{1,1}, \dots, Y_{1,n_1}$  and  $Y_{2,1}, \dots, Y_{2,n_2}$  have been defined for the responses in this problem. These responses are strictly positive numbers, and an assumption of independence is reasonable. Formulate a two-sample model using gamma distributions. For one group, write the form of the log likelihood that will need to be computed to find estimates and other inferential quantities.

**Answer**

Assuming the random variables defined as given, taking observed values that are strictly positive and iid.

Each group is then modeled with a (potentially different) Gamma distribution parameterized by the  $(\alpha, \beta)$  parameters (shape and rate, respectively):

$$Y_{g,i} \stackrel{\text{iid}}{\sim} \text{Gamma}(\alpha_g, \beta_g), \quad g \in \{1, 2\}, \quad \text{and } i = 1, \dots, n_g$$

Where, for our purposes  $n_1 = n_2$  (equal sample sizes between the two groups of Gamma-distributed random variables).

With pdf of the form:

$$f(y | \alpha_g, \beta_g) = \frac{\beta_g^{\alpha_g}}{\Gamma(\alpha_g)} y^{\alpha_g-1} e^{-\beta_g y}, \quad y > 0, \alpha_g > 0, \beta_g > 0$$

Using the pdf, we then take the log to define a single group  $g$ 's log-likelihood function by:

$$\ell_g(\alpha_g, \beta_g) = \sum_{i=1}^{n_g} \log f(y_{g,i} | \alpha_g, \beta_g) = n_g \left( \alpha_g \log \beta_g - \log \Gamma(\alpha_g) \right) + (\alpha_g - 1) \sum_{i=1}^{n_g} \log y_{g,i} - \beta_g \sum_{i=1}^{n_g} y_{g,i}$$

Note: The log-likelihood function is defined by 2 sufficient statistics, of the form:

$$T_{g,1} = \sum_{i=1}^{n_g} \log y_{g,i} \quad \text{and} \quad T_{g,2} = \sum_{i=1}^{n_g} y_{g,i}$$

So, for the score equations, computing MLEs (say, via Newton–Raphson or some other optimization methods) is done via:

$$\frac{\partial \ell_g}{\partial \alpha_g} = n_g \log \beta_g - n_g \psi_0(\alpha_g) + T_{g,1}, \quad \frac{\partial \ell_g}{\partial \beta_g} = \frac{n_g \alpha_g}{\beta_g} - T_{g,2}$$

With the Hessian given by:

$$\frac{\partial^2 \ell_g}{\partial \alpha_g^2} = -n_g \psi_1(\alpha_g), \quad \frac{\partial^2 \ell_g}{\partial \beta_g^2} = -\frac{n_g \alpha_g}{\beta_g^2}, \quad \frac{\partial^2 \ell_g}{\partial \alpha_g \partial \beta_g} = \frac{n_g}{\beta_g}$$

Giving the Hessian:

$$\begin{bmatrix} \frac{\partial^2 \ell_g}{\partial \alpha_g^2} & \frac{\partial^2 \ell_g}{\partial \alpha_g \partial \beta_g} \\ \frac{\partial^2 \ell_g}{\partial \beta_g \partial \alpha_g} & \frac{\partial^2 \ell_g}{\partial \beta_g^2} \end{bmatrix} = \begin{bmatrix} -n_g \psi_1(\alpha_g) & \frac{n_g}{\beta_g} \\ \frac{n_g}{\beta_g} & -\frac{n_g \alpha_g}{\beta_g^2} \end{bmatrix}$$

where  $\psi_0(\cdot)$  and  $\psi_1(\cdot)$  are the digamma and trigamma functions, respectively (following notation convention seen on Wikipedia).

Then, taking the individual group log-likelihoods, we calculate the full two-sample log-likelihood as the sum (again, noting iid assumption between and within groups):

$$\begin{aligned}
\ell(\alpha_1, \beta_1, \alpha_2, \beta_2) &= \ell_1(\alpha_1, \beta_1) + \ell_2(\alpha_2, \beta_2) \\
&= \sum_{g=1}^2 \sum_{i=1}^{n_g} \log f(y_{g,i} \mid \alpha_g, \beta_g) \\
&= n_1 \left( \alpha_1 \log \beta_1 - \log \Gamma(\alpha_1) \right) + (\alpha_1 - 1) \sum_{i=1}^{n_1} \log y_{1,i} - \beta_1 \sum_{i=1}^{n_1} y_{1,i} \\
&\quad + n_2 \left( \alpha_2 \log \beta_2 - \log \Gamma(\alpha_2) \right) + (\alpha_2 - 1) \sum_{j=1}^{n_2} \log y_{2,j} - \beta_2 \sum_{j=1}^{n_2} y_{2,j}
\end{aligned}$$

## 2.

Find maximum likelihood estimates and 95% Wald theory intervals for the parameters of each group. Recall that, in the data file, the first column of values is Group 1 and the second column of values is Group 2.

### Answer

As noted in part 1)., the log-likelihood is of the form:

$$\ell_g(\alpha_g, \beta_g) = n_g(\alpha_g \log \beta_g - \log \Gamma(\alpha_g)) + (\alpha_g - 1) \sum_{i=1}^{n_g} \log Y_{g,i} - \beta_g \sum_{i=1}^{n_g} Y_{g,i} \quad \text{where } g \in \{1, 2\}$$

Setting the score functions to zero gives the standard MLE system of equations.

The Fisher Information (negative of the Hessian of the log-likelihood evaluated at the MLE,  $(\alpha_g, \beta_g)$ ) then is:

$$I_g(\alpha_g, \beta_g) = \begin{pmatrix} n_g \psi_1(\alpha_g) & -n_g/\beta_g \\ -n_g/\beta_g & n_g \alpha_g / \beta_g^2 \end{pmatrix}$$

Taking these quantities, the Wald covariance is then given by:

$$\widehat{\text{Var}} \begin{pmatrix} \hat{\alpha}_g \\ \hat{\beta}_g \end{pmatrix} = I_g(\hat{\alpha}_g, \hat{\beta}_g)^{-1}$$

After numeric approximation, taking square roots where appropriate (square root of the variance is SE), and evaluating the typical expression for confidence intervals, we have (with each group having  $n_1 = n_2 = 50$  samples):

Group	$\hat{\alpha}$	SE( $\hat{\alpha}$ )	95% CI for $\alpha$	$\hat{\beta}$	SE( $\hat{\beta}$ )	95% CI for $\beta$
1	3.497	0.669	(2.186, 4.808)	1.519	0.312	(0.907, 2.131)
2	1.626	0.298	(1.042, 2.210)	0.726	0.155	(0.421, 1.031)

Note, to be explicit about the formula for confidence intervals:

$$\exp(\log \hat{\alpha} \pm z_{1-\gamma/2} \text{SE}(\log \hat{\alpha})) \quad \text{and} \quad \exp(\log \hat{\beta} \pm z_{1-\gamma/2} \text{SE}(\log \hat{\beta}))$$

Where  $\gamma = 0.05, 1 - \frac{\gamma}{2} = 0.975$

And to be explicit about the optimization method used: Using R's `optim` function for maximization (minimize negative log-likelihood), which is a “quasi-Newton method”, i.e., using the original log-likelihood, first derivative, and second derivative (also using `method = 'BFGS'`).

### 3.

Using a likelihood ratio test, determine whether you would reject a model having a common gamma distribution for both groups in favor of a model having separate gamma distributions for each of the two groups. Produce a plot of the estimated densities for each group (both densities on the same plot).

#### Answer

Define the (nested) hypotheses by:

- $H_0 : \alpha_1 = \alpha_2$  and  $\beta_1 = \beta_2$  (one common Gamma for both groups, only 2 unique parameters between groups).
- $H_1 : (\alpha_1 \neq \alpha_2)$  and  $(\beta_1 \neq \beta_2)$  (two separate Gammas, 4 unique parameters between groups).

Let  $\hat{\theta}_0 = (\hat{\alpha}_0, \hat{\beta}_0)$  be the MLE under  $H_0$  (fitted using pooled data), and  $\hat{\theta}_1 = ((\hat{\alpha}_1, \hat{\beta}_1), (\hat{\alpha}_2, \hat{\beta}_2))$  the MLEs fitted to each group separately.

The LRT statistic is of the form:

$$\Lambda = -2\{\ell(\hat{\theta}_0) - \ell(\hat{\theta}_1)\} \xrightarrow{d} \chi_2^2$$

With degrees of freedom 2 from (full - reduced = 4 - 2), i.e.,  $H_1$  has two more “free” parameters than  $H_0$ .

Using the same optimization method using in part 2., we calculate:

- Separate-group MLEs:  $\hat{\alpha}_1 = 3.497$ ,  $\hat{\beta}_1 = 1.519$ , and  $\hat{\alpha}_2 = 1.626$ ,  $\hat{\beta}_2 = 0.726$
- Common (pooled) MLEs:  $\hat{\alpha}_0 = 2.202$ ,  $\hat{\beta}_0 = 0.970$

Where:

$$\ell(\hat{\theta}_1) = -163.447 (= \ell_1(\alpha_1, \beta_1) + \ell_2(\alpha_2, \beta_2) = -76.22814 + -87.21866), \quad \ell(\hat{\theta}_0) = -167.526$$

Using the log-likelihood values above, the LRT statistic is:

$$\Lambda = -2(-167.526 - (-163.447)) = 8.158$$

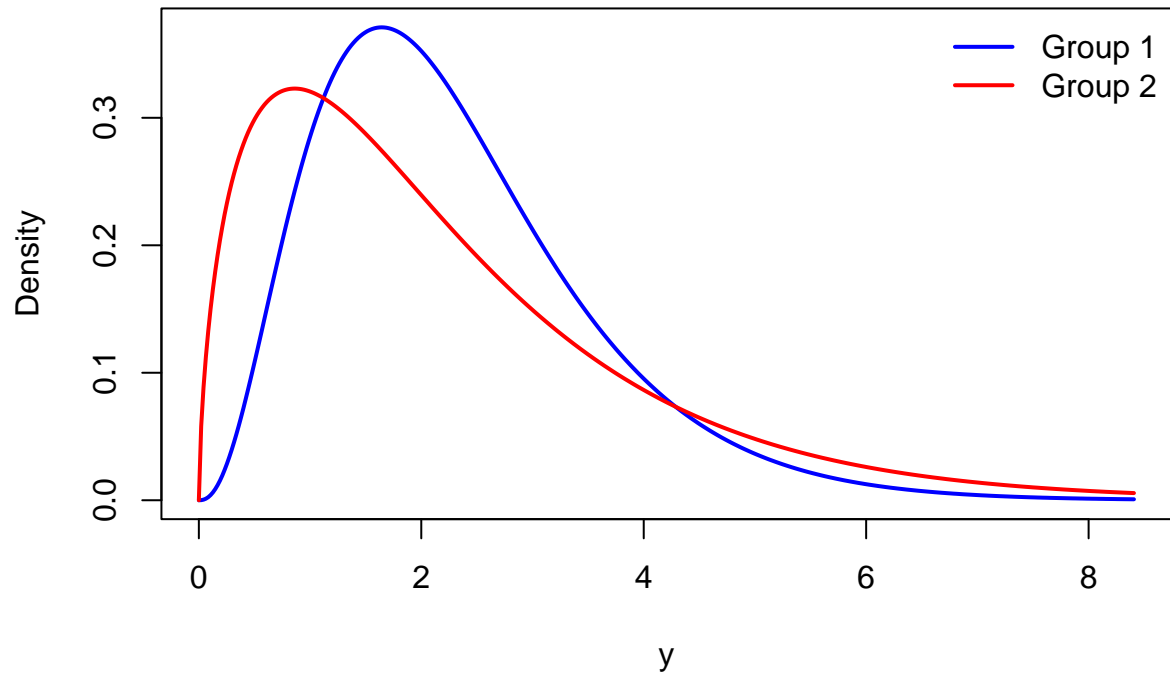
With corresponding p-value (with reference distribution  $\chi_2^2$ ):

$$p\text{-value} = 0.01693$$

Reducing the question to an “Accept”/“Reject” framework, we Reject  $H_0$  at the  $\alpha = 0.05$  level (or make a “strength of evidence” argument to say we have strong evidence in favor of rejecting the null hypothesis). Interpreting this, we’d say that modeling the two groups as separate (differently parametrized) Gamma distributions seems a better fit than pooling them together as a single Gamma distribution (with shared shape and rate parameters).

And a graph!

**MLE Gamma Densities by Group**



4.

Find maximum likelihood estimates and 95% Wald theory intervals for the expected value of each group. Also produce a 95% interval for the difference in expected values (Group 1 minus Group 2).

**Answer**

For a Gamma distributed random variable  $Y$ , whose distribution is parametrized by the shape and rate parameters, the expected value is of the form:

$$E[Y] = \frac{\alpha}{\beta}$$

For our case, thus the expected values for each group are given by:

$$\mu_g = \frac{\alpha_g}{\beta_g}, \quad g = 1, 2$$

Let  $(\hat{\alpha}_g, \hat{\beta}_g)$  be the MLEs, then:

$$\hat{\mu}_g = \frac{\hat{\alpha}_g}{\hat{\beta}_g}, \quad g = 1, 2$$

Note then that the covariance matrix of the MLEs (the inverse observed information) is given by:

$$\hat{\Sigma}_g = \begin{bmatrix} \widehat{\text{Var}}(\hat{\alpha}_g) & \widehat{\text{Cov}}(\hat{\alpha}_g, \hat{\beta}_g) \\ \widehat{\text{Cov}}(\hat{\alpha}_g, \hat{\beta}_g) & \widehat{\text{Var}}(\hat{\beta}_g) \end{bmatrix}$$

Then, using the delta method with gradient given by:

$$\begin{aligned} \nabla \mu_g(\alpha, \beta) &= \left( \frac{\partial}{\partial \alpha}(\mu_g), \frac{\partial}{\partial \beta}(\mu_g) \right) \\ &= \left( \frac{\partial}{\partial \alpha} \left( \frac{\alpha_g}{\beta_g} \right), \frac{\partial}{\partial \beta} \left( \frac{\alpha_g}{\beta_g} \right) \right) \\ &= \left( \frac{1}{\beta}, -\frac{\alpha}{\beta^2} \right) \end{aligned}$$

Then

$$\widehat{\text{Var}}(\hat{\mu}_g) = \nabla \mu_g^\top \hat{\Sigma}_g \nabla \mu_g$$

Taking the SE then is just taking the root of the above quantity.

Note: When calculating the quantities of interest, we use the sample (observed) variances of  $\hat{\alpha}$  and  $\hat{\beta}$ .

Then, we may construct a 95% Wald interval for  $\mu_g$  by:

$$\hat{\mu}_g \pm z_{1-\frac{\gamma}{2}} \text{SE}(\hat{\mu}_g) = \hat{\mu}_g \pm 1.96 \text{SE}(\hat{\mu}_g)$$

(Using 1.96 based on the Standard Normal distribution, where  $\gamma = 0.05$ , as Wald Theory uses asymptotic results.)

For the difference  $\mu_1 - \mu_2$ , treat groups as independent, i.e.,  $\text{Cov}(\hat{\mu}_1, \hat{\mu}_2) = 0$ , so

$$\begin{aligned}\widehat{\text{Var}}(\hat{\mu}_1 - \hat{\mu}_2) &= \widehat{\text{Var}}(\hat{\mu}_1) + \widehat{\text{Var}}(\hat{\mu}_2) + 2\widehat{\text{Cov}}(\hat{\mu}_1, \hat{\mu}_2) \\ &= \widehat{\text{Var}}(\hat{\mu}_1) + \widehat{\text{Var}}(\hat{\mu}_2)\end{aligned}$$

Taken together, the quantities of interest we estimate are:

Table 2: Gamma means and 95% Wald intervals

group	mu_hat	se	ci_L	ci_U
group1	2.302	0.174	1.961	2.644
group2	2.240	0.248	1.753	2.726
difference (1 - 2)	0.063	0.303	-0.532	0.657



## 5.

Test whether the two groups should be considered significantly different using a two-sample  $t$ -test. (Take square roots if you think it makes the data look more symmetric for each group, though this is optional.) Does your result agree with the likelihood ratio test? Does it agree with the interval for difference in expected values?

### Answer

The observed group sample means are given by:

- Group 1:  $\bar{y}_1 \approx 2.29$  (2.30 expected),
- Group 2:  $\bar{y}_2 \approx 2.24$  (2.24 expected).

Also, the observed sample standard deviation is given by.

- Group 1:  $s_1 \approx 1.31$
- Group 2:  $s_2 \approx 1.75$
- Pooled:  $s_0 \approx 1.55$

And we have equal number of samples per group,  $n_1 = n_2 = 50$ , which taken together could justify use of the two-sample  $t$ -test.

The hypotheses we may then test are of the form:

$$H_0 : \mu_1 = \mu_2 \quad \text{vs.} \quad H_A : \mu_1 \neq \mu_2$$

Since there are differences in variance between Group 1 and Group 2, I used the Welch two-sample  $t$ -test (which does not assume equal variances between groups). Notably, while I cannot emphatically justify using Welch, I prefer to use it in this instance because it reduces to the (“pooled”) two-sample  $t$ -test when the sample variances are close (which I think they are, though still distinct between groups).

Additionally: While the formulas below are for the Welch two-sample  $t$ -test, I also calculate and provide formulas for the (non-Welch, standard) two-sample  $t$ -test.

That being said, the relevant formulas for the Welch are of the form:

$$t_\nu = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \approx 0.15$$

Where the degrees of freedom are calculated using the Welch-Satterthwaite approximation:

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

The corresponding p-value is then:  $p \approx 0.84$  (using Welch, and with no transformation).

If we repeat this process *after* doing a **sqrt** transformation on the observed samples, the result is very similar: The test statistic is  $t_{98} \approx 0.78$  with corresponding  $p \approx 0.44$  (again, using Welch). So, reducing to a decision “Accept”/“Reject” or even a “strength of evidence” argument, the interpretation remains largely the same.

These results may (at first) seem opposed to the Likelihood ratio test from part 3), but it isn’t. Importantly, these two methods (LRT test vs. the two-sample t-test) correspond to different tests. The two-sample  $t$ -test does not provide evidence of a significant difference in means between the two groups. This is also consistent with the Wald interval for the difference in means  $\mu_1 - \mu_2$ , containing 0. But LRT is not a direct comparison of means; instead, the LRT tests whether the two *distributions* (particularly, the parameters and distributional form, so encompassing distributional shape, spread, and other aspects of the distributional form) differ significantly between groups.

Also, the results of the two-sample test are consistent when we use the normal “pooled” two-sample t-test, and also when we transform the original data using the **sqrt** transformation. All the relevant statistics and summary information of the calculations considered are in the following table:

Table 3: Table of Two-Sample Test Statistics

analysis	mean1	mean2	sd1	sd2	t_stat	df	p_value	ci_lower	ci_upper
Raw, Welch	2.302	2.240	1.308	1.755	0.203	90.608	0.840	-0.552	0.678
Sqrt, Welch	1.463	1.386	0.406	0.571	0.781	88.471	0.437	-0.119	0.274
Raw, pooled	2.302	2.240	1.308	1.755	0.203	98.000	0.840	-0.552	0.677
Sqrt, pooled	1.463	1.386	0.406	0.571	0.781	98.000	0.436	-0.119	0.274

Misc: There are two notes to the above. First is, the Welch seemed most appropriate given differences in variance (and differences in shape-rate parameters), hence why it was formally noted and used. However, since the two-sample t-test was also calculated, the formulas used are as follows:

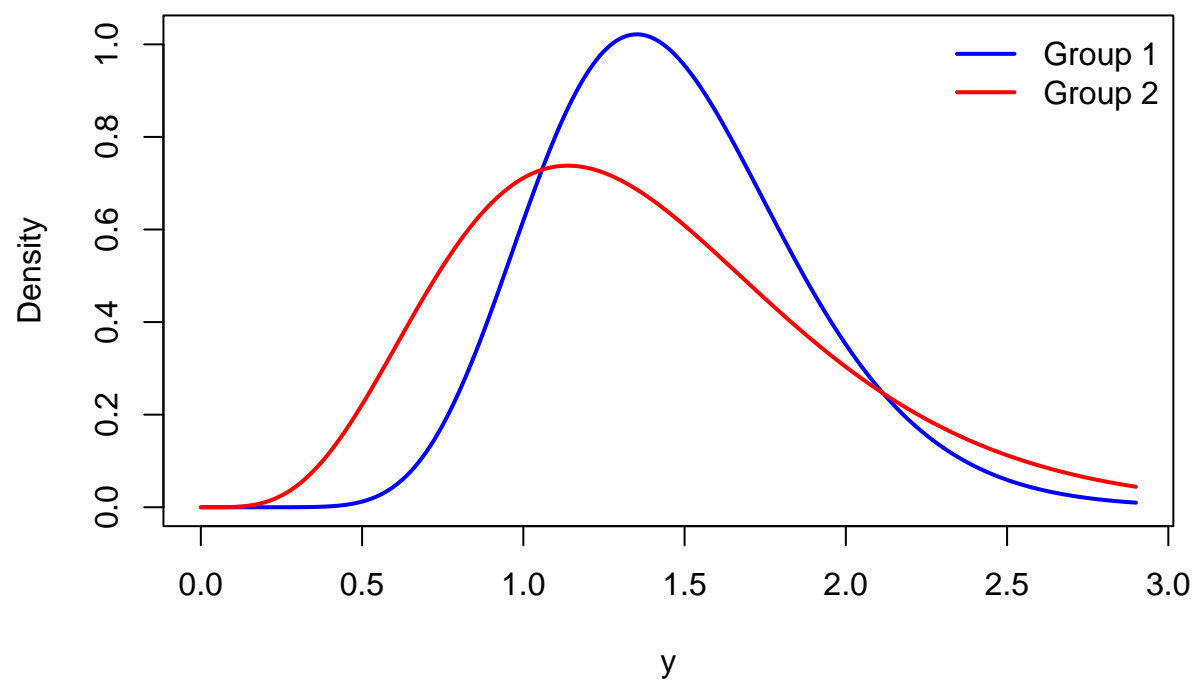
$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$t_{df} = \frac{\bar{y}_1 - \bar{y}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad df = n_1 + n_2 - 2$$

Where  $df$ ,  $s_p^2$ , and  $t_P$  as given in the above table.

Also, since I included the **sqrt** evaluation as well in the table, below is visual evidence that the distributions appear more symmetric after **sqrt** transformation to the data (which also gives further visual evidence of the spread/variance being different between the two groups).

**Sqrt -- MLE Gamma Densities by Group**



## 6.

Find maximum likelihood estimates and 95% Wald theory intervals for the mode of each group. Also produce a 95% interval for the difference in modes (Group 1 minus Group 2).

### Answer

For  $Y$  a Gamma-distributed random variable, parametrized in shape–rate form, the mode is of the form:

$$\text{mode} = \begin{cases} \frac{\alpha - 1}{\beta}, & \alpha > 1, \\ 0, & 0 < \alpha \leq 1 \end{cases}$$

As the MLEs of  $\hat{\alpha}_1, \hat{\alpha}_2 > 1$ , we won't worry about the second condition of the mode, i.e., where  $0 < \alpha \leq 1$ .

As defined previously,  $(\hat{\alpha}_g, \hat{\beta}_g)$  are the MLEs for group  $g$  with covariance given by  $\hat{\Sigma}_g$ .

For  $\hat{\alpha}_g > 1$ , we use the delta method with the expressions:

$$m_g(\alpha, \beta) = \frac{\alpha - 1}{\beta}, \quad \nabla m_g(\alpha, \beta) = \left( \frac{1}{\beta}, -\frac{\alpha - 1}{\beta^2} \right)$$

Again, mirroring the formulae used in part 4).

We then have the quantities:

$$\widehat{\text{Var}}(\hat{m}_g) = \nabla m_g^\top \hat{\Sigma}_g \nabla m_g, \quad \text{SE}(\hat{m}_g) = \sqrt{\widehat{\text{Var}}(\hat{m}_g)}$$

To construct a 95% Wald CI, given by:

$$\hat{m}_g \pm z_{1-\frac{\gamma}{2}} \text{SE}(\hat{m}_g) = \hat{m}_g \pm 1.96 \text{SE}(\hat{m}_g)$$

For the difference  $m_1 - m_2$ , independence of groups gives

$$\begin{aligned} \widehat{\text{Var}}(\hat{m}_1 - \hat{m}_2) &= \widehat{\text{Var}}(\hat{m}_1) + \widehat{\text{Var}}(\hat{m}_2) + 2\widehat{\text{Cov}}(\hat{m}_1, \hat{m}_2) \\ &= \widehat{\text{Var}}(\hat{m}_1) + \widehat{\text{Var}}(\hat{m}_2) \end{aligned}$$

Calculating these quantities gives us the following table:

Table 4: Gamma modes and 95% Wald intervals

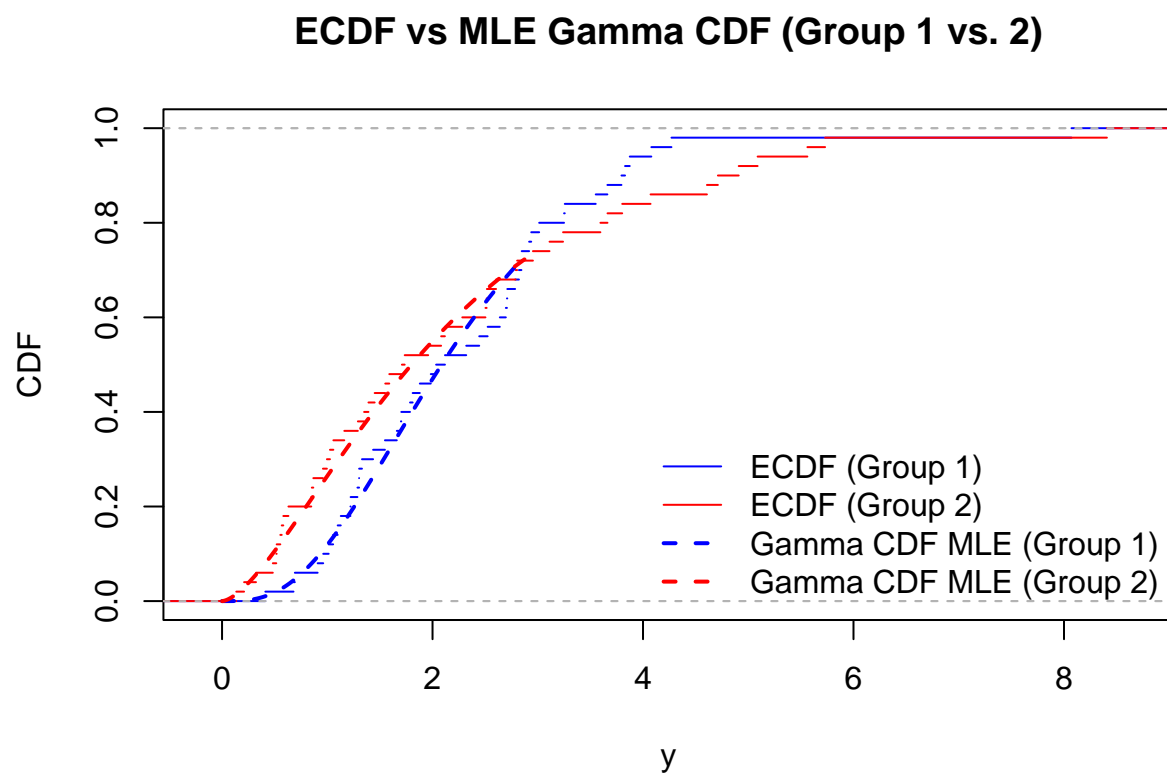
group	mode_hat	se	ci_L	ci_U
group1	1.644	0.177	1.297	1.991
group2	0.862	0.270	0.334	1.391
difference (1 - 2)	0.782	0.323	0.149	1.414

7.

Although model assessment has not yet been covered formally, it is intuitive that the estimated distribution function (CDF) under our model and the empirical distribution function of the data should be similar. Produce plots of the estimated distribution function for each group with the empirical distribution function overlaid.

### Answer

To avoid any potential “code dump” regarding this question, I just wanted to note that the **stats** package contains an **ecdf** function. This was used for the following comparisons with the “expected” line (based on the MLEs  $\hat{\alpha}_g, \hat{\beta}_g, g \in \{1, 2\}$ ).



8.

Write a short paragraph giving your conclusions about this group comparison.

**Answer**

Comparing the ECDFs to their fitted Gamma distributions, both groups match reasonably well, although Group 1 has more visually apparent differences in the 2–4 range (domain). In terms of expectation/central tendency, the two groups have very similar means, as evidenced when comparing means via the two-sample t-tests. It is then perhaps unsurprising that the Wald confidence interval for the difference in expected values includes zero. However, the likelihood ratio test provides some evidence that the overall distributions differ, with Group 1 and Group 2 having distinct shape–rate parameter combinations (their individual parameter values). This is also evidenced somewhat in the above visual, as the ECDFs and fitted Gamma CDFs look different between the two groups. To put it hopefully succinctly: The results from the prior parts (1 through 7) suggest that while the groups do not differ significantly in their means, their distributional shapes are different enough to warrant modeling them separately.