EMPIRY MEETS THEORY: ON MAXIMUM LIKELIHOOD ESTIMATION FOR SURVIVAL TIME OF GUINEA PIGS.

ANDRÉ THÉO, **GANESH** GAUTHAM AND **MISHRA** PIYUSH

Turing Centre for Living Systems & Aix-Marseille Université

Contents

1. Introduction

The gamma distrbution is often used to study and sketch the evolution of waiting times in everyday applications such as internet connectivity and traffic signals. In the past few decades, they have also been used to model such waiting times in biological systems and experiments. Here, we utilise survival time data from Bjerkedal (1960) [2], who studied the guinea pig's capacity to resist infection by tubercle bacilli bacteria using a quantitative approach. Since the guinea pigs were administered with different doses of infectious bacteria, arranging the former according to increasing survival time is highly indicative of their antibiotic resistance capacity.

In this report, we shall walk the reader through our search for compatible distributions and subsequently verify the suitability of the best model. We then perform a complete theoretical analysis of the gamma distribution (which turned out to be the best fit) and describe the Maximum-Likelihood equations and estimators. We then conclude with a visualisation of the evolution of the mean-squared error for different Maximum-Likelihood estimators (MLE) used for the parameters of the gamma distribution.

All the code used for simulations and the generation of various plots in this report are available in the following repository.

2. Establishing the statistical model

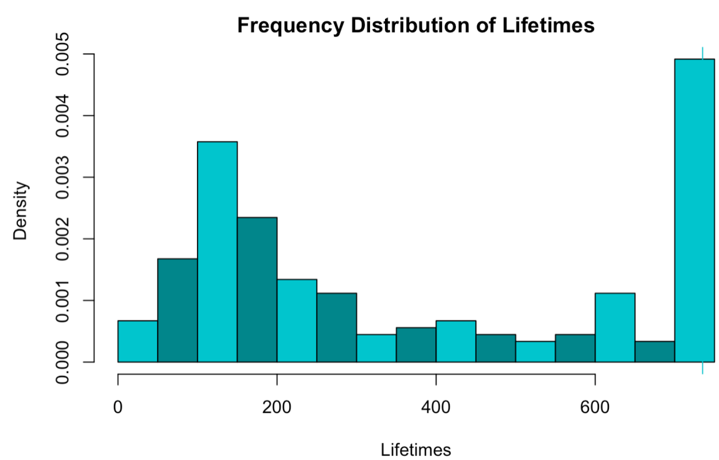

Figure 1. Normalised distribution of lifetimes: a censored data gives us two peaks

2.1. Basis.

A whole world of distribution we could use to fit our data just opened to us. First, notice that the data set has been subject to censorship on the right (this justifies the existence of a peak on the right). A good way to approximate our data, to account for this phenomenon, would be to fit a bimodal distribution. Rather, let us start simple and ignore this rightmost peak at the moment. Instead, try to fit a unimodal distribution instead. Here we have several choices: Normal distribution? Weibull's distribution? Laplace's distribution? Each of these options is likely with some plausibility.

2.2. Finding Compatible Distributions.

The key principle in this section is to be able to find a distribution that fits well with the distribution of guinea pig lifetimes. A naive approach here would be to first observe the different compatible distributions with data that is uncensored. So, we first assume that censoring does not provide any information. This distribution is exhibited in Fig. 2

Just by looking at the histogram, it is very evident that this distribution does not follow the normal law. As the section progresses, we develop why it isn't the normal law, and what other distributions could be tested.

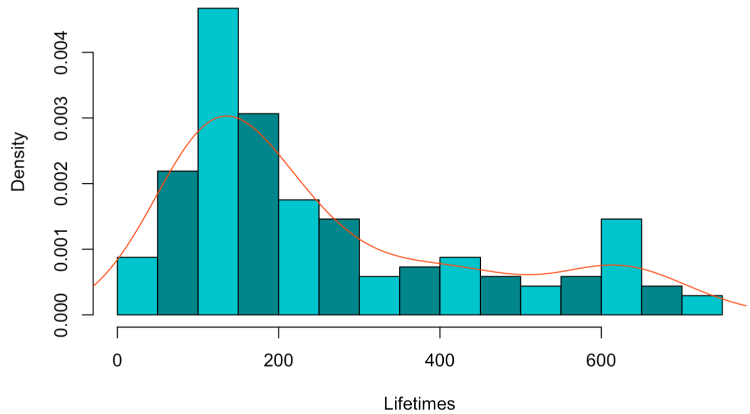

FIGURE 2. Normalised distribution and density curve of the observed data on removing the censored data

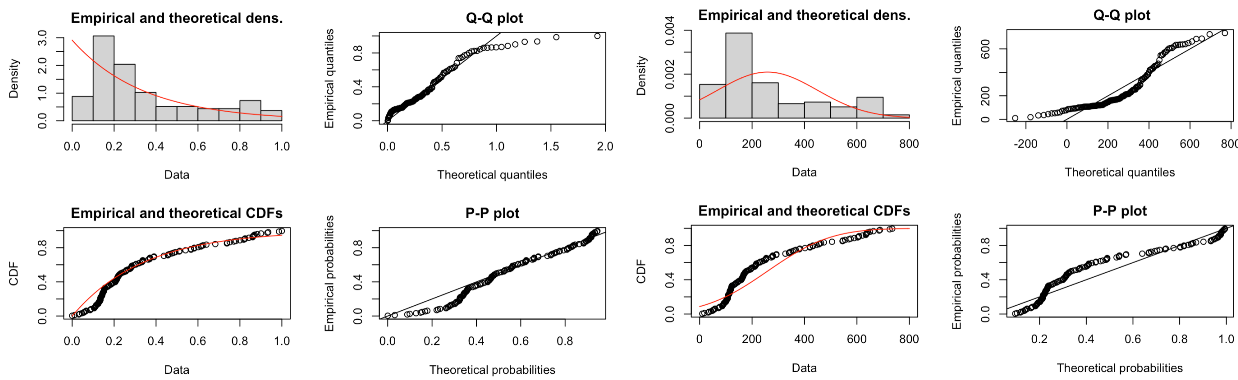

FIGURE 3. Compatibility comparison of the distribution with exponential (left) and normal (right) laws by looking at the empirical and theoretical densities, cumulative distributions, Q-Q plots and P-P plots

In Fig. 3 we check fitted distributions of exponential and normal laws to understand which works better with the empirical data. We observe that we are fairly able to fit the laws, these aren't probably the best laws to fit the distribution with.

We, thus, make use of the Cullen and Frey method to make sense of the available distribution, as shown in Fig. 4. A Cullen and Frey plot [3] makes use of kurtosis and skewness of a data to understand the closest compatible distribution with the empirical data. Skewness and kurtosis are the third and the fourth moments of standardised variables respectively – skewness is the degree of asymmetry observed in a probability distribution while kurtosis givesd us information on the degree of presence of outliers in a distribution. We see that there is a higher kurtosis for the empirical data than the gamma distribution, and a higher skewness than the uniform and the normal distributions. It can be observed that there is indeed more skewness in the empirical data than in the fitted normal distribution, as was also evident from Fig 3. So we can successfully eliminate normal and exponential distributions. From the Cullen and Frey plot, it can be observed that the Gamma distribution seems to be the closest to the given distribution.

We thus fit the Gamma distribution on the data as shown in Fig. 5.

2.3. **Is this inference any good?** While we see that the Gamma distribution fits “well" with the observed data, we can't really eyeball for statistical significance. A good question to ask here is, “How can the compatibility of a Gamma distribution be confirmed?". A good starting point for answering this question is the **Kolmogorov-Smirnov test**. The Kolmogorov-Smirnov test is a non-parametric hypothesis test that measures the probability that a chosen dataset is drawn from the same parent population as a continuous model, of course, for a one-sample Kolmogorov-Smirnov test. The null hypothesis of this test, thus, states that

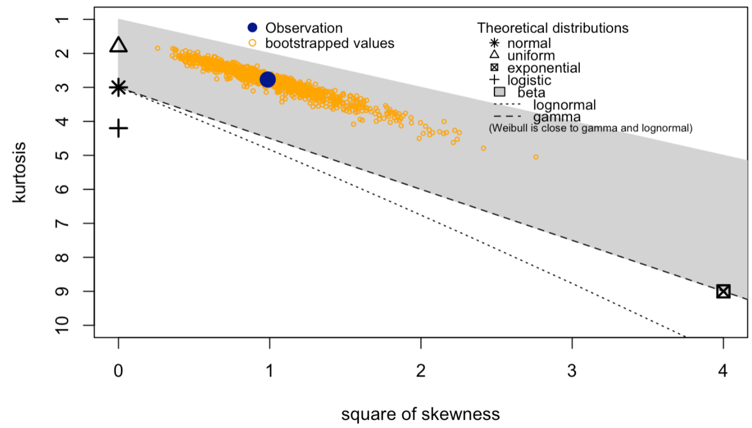

FIGURE 4. Cullen and Frey graph of the uncensored data: notice that the distribution is closer to the uniform and the normal distributions than to the exponential distribution

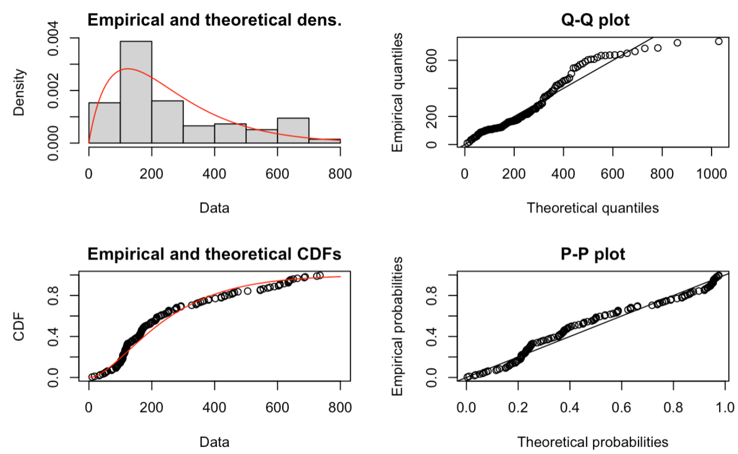

FIGURE 5. Compatibility of Gamma distribution by looking at empirical and theoretical densities, cumulative distributions, Q-Q plots and P-P plots

the true distribution is equal to the hypothesised distribution. When we, thusly, carry out the Kolmogorov-Smirnov test against the Gamma distribution, we get a p-value of 2.2×10^{-16}, which is significantly lower than an α of 0.05 for a 95% confidence. This means, that we have to reject the null hypothesis, which further means that the distribution does not follow the Gamma law! So, do the results that we infer from our plots and the Kolmogorov-Smirnov test give a different picture?

While it might seem to be the best option to "confirm" the hypothesis using the Kolmogorov-Smirnov test, there are some caveats – since this test does not take into account the uncertainty of estimation, the p-value obtained from this test might not be the best source of reliability for us. Thus, unfortunately for us, we cannot just fit a distribution and carry out the test on estimated parameters. One manoeuvre for this could be that we could use another normality test, the Lilliefors test which is a modified test on Kolmogorov-Smirnov test that takes into account the uncertainty of estimation.

The underlying theory of Kolmogorov-Smirnov test requires the two curves in question to be independent of each other, i.e., the model must be derived from another dataset, or from external considerations for us to be able to apply the Kolmogorov-Smirnov test. The Kolmogorov-Smirnov statistic, however, can still be calculated, and the significance level of the difference between the EDF curves can be assessed using bootstrap resamples of the original dataset. Bootstrap resampling is easy to understand and implement, and the theory behind it ensures that the final significance levels are impartial in a variety of situations, as shown in Fig. 6.

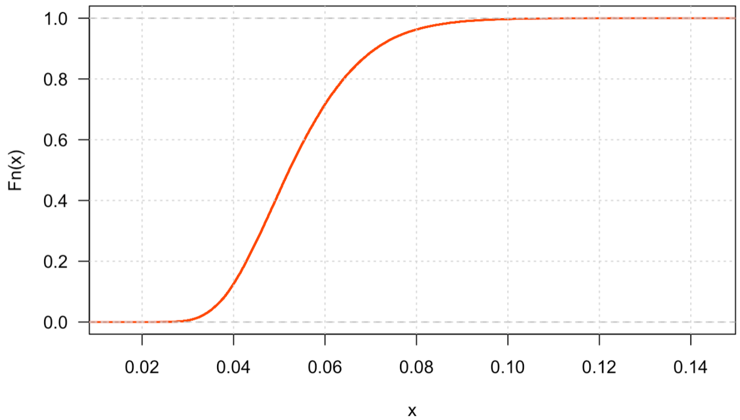

FIGURE 6. Empirical cumulative distribution of Kolmogorov-Smirnov test statistic simulations by bootstrapping

With bootstrapping we get a p-value of 0.004, which is inching towards significance. Why is it still not statistically significant? A sample will never really follow a specific distribution exactly. The goal is not to determine, with certainty, what distribution the data follows, but to argue which distribution is the most compatible.

2.4. **Best model criterion.** Since statistical tests aren't really reliable for us to fully confirm the functioning of the different statistical models that we are taking into consideration, we have to rely on some other parameters for confirming that. We, thus, use certain "goodness-of-fit" parameters that would compare and contrast between the different models taken into consideration.

	Normal	Exponential	Gamma	Log Normal
Kolmogorov-Smirnov statistic	0.1801772	0.1951092	0.09777311	0.06397839
Cramer-von Mises statistic	1.2835690	0.9043416	0.35741039	0.14116596
Anderson-Darling statistic	7.3342075	5.7268612	2.11715764	1.14846129
Akaike's Information Criterion	1831.232	1798.393	**1772.009**	1774.691
Bayesian Information Criterion	1837.072	1801.313	**1777.849**	1780.531

We see that Akaike's information criterion and the Bayesian information criterion for the Gamma distribution is the lowest among all the distributions employed, which suggests that it is indeed the best among all the distributions that fits the empirical data. With regard to what has been done thus far, let us assume the following

Modality of the distribution: The arguments presented have been considering the entire distribution to be unimodal. This can be justified since it is assumed that the censored data does not provide any information. Further, we carry out the Bajgier and Aggarwal test [1] to further confirm this presumption.

Claim. *The sample of lifetimes* $\boldsymbol{X} = (x_1, x_2, \ldots, x_n)$ *has been sampled from a random vector of i.i.d. variables* $(X_1, X_2, ..., X_n)$ *with* X_1 *following a gamma distribution* $\boldsymbol{\mathcal{G}}(\alpha_0, \gamma_0)$ *where the parameters* α_0, γ_0 *are elements of the parameters space* $\boldsymbol{\Theta} \stackrel{set}{=} \mathbb{R}_+^* \times \mathbb{R}_+^*$. *Let us denote by* $f_{\boldsymbol{X}}(\cdot\ ;\ \alpha, \gamma)$ *the density of the gamma distribution* $\boldsymbol{\mathcal{G}}(\alpha, \gamma)$. *Then the model* $\mathfrak{M} = \{f_{\boldsymbol{X}}(\cdot\ ;\ \alpha, \gamma),\ (\alpha, \gamma) \in \boldsymbol{\Theta}\}$ *defines a parametric statistical model.*

Proof. Here we briefly apply ourselves to discuss why is this true. First notice that the sampling hypothesis *(H-S)* is satisfied by the claim's assumptions. Next, the choice of $\boldsymbol{\Theta}$ allows us to correctly define the family

$$\mathcal{F} = \left\{ \boldsymbol{F_\theta}(x) \overset{\textbf{set}}{=\!=} \int_{[0,x]} f_{\boldsymbol{X}}(y;\boldsymbol{\theta})dy \;\middle|\; \boldsymbol{\theta} \in \boldsymbol{\Theta} \right\}$$

so that *(H-P)* is satisfied.

Remark: *One could notice that* $\mathfrak{M}$ *also satisfies (H-Id) and (H-D) with our choice of* $\boldsymbol{\Theta}$ *and finally it happens that under our assumptions, the true parameter* (α_0, γ_0) *is an element of* $\boldsymbol{\Theta}$*. thus, not only* $\mathfrak{M}$ *is a parametrical statistical model, but it is also a well-defined model. An even more powerful result, since our model trivially satisfies all hypotheses* ***(H1-6)****,* $\mathfrak{M}$ *is also a regular statistical model. The last result will implicitly used a lot during the MLE part.* **QED**

3. Generalities on the Gamma distribution

Computing the first moments of a variable following some distribution is always a good habit. Not only it is useful to get a taste of what the distribution will look like, but it also will be of great use when working with estimators of our parameters. From our assumptions, suppose each $X_i \sim \boldsymbol{\mathcal{G}}(\alpha, \gamma)$ and let us compute its first two moments. Recall X_i admits the following probability density function (p.d.f.)

$$f\left(X_i \,|\, \boldsymbol{\theta}\right) = f_{\boldsymbol{X}}(x\,;\alpha,\gamma) = \frac{1}{\boldsymbol{\Gamma}(\alpha)\gamma^\alpha} x^{\alpha-1} e^{-x/\gamma} \cdot \mathbf{1}_{\mathbb{R}^+}(x)$$

With $\boldsymbol{\theta} = (\alpha, \gamma)$, hence

$$\mathbb{E}\left[\boldsymbol{X}\right] = \int_{\mathbb{R}} x \cdot f_{\boldsymbol{X}}(x;\alpha,\gamma)dx = \frac{1}{\boldsymbol{\Gamma}(\alpha)\gamma^\alpha} \int_{\mathbb{R}^+} x^{(\alpha+1)-1} e^{-x/\gamma} dx$$

the change of variable $x = \gamma y \to dx = \gamma dy$ yields

$$\mathbb{E}\left[\boldsymbol{X}\right] = \frac{\gamma}{\boldsymbol{\Gamma}(\alpha)\gamma^\alpha} \int_{\mathbb{R}^+} (\gamma y)^{(\alpha+1)-1} e^{-y} dy = \frac{\gamma}{\boldsymbol{\Gamma}(\alpha)} \boldsymbol{\Gamma}(\alpha+1) = \gamma\alpha$$

notice we used the fact that $\boldsymbol{\Gamma}(z) = (z-1)\boldsymbol{\Gamma}(z-1)$ provided $z \in \mathbb{C}$ and $\mathfrak{Re}(z) > 0$. Moving on to the variance $\mathbb{V}[\boldsymbol{X}]$, computations are of the same flavor, that is to say

$$\mathbb{V}\left[\boldsymbol{X}\right] = \mathbb{E}\left[\boldsymbol{X}^2\right] - \mathbb{E}\left[\boldsymbol{X}\right]^2 = \frac{1}{\boldsymbol{\Gamma}(\alpha)\gamma^\alpha} \int_{\mathbb{R}^+} x^{(\alpha+2)-1} e^{-x/\gamma} dx - \gamma^2\alpha^2$$

Once again, substitute $x = \gamma y \to dx = \gamma dy$ to reach

$$\mathbb{V}[\boldsymbol{X}] = \gamma^2 \left[\frac{\boldsymbol{\Gamma}(\alpha+2)}{\boldsymbol{\Gamma}(\alpha)} - \alpha^2 \right] = \gamma^2 \left[(\alpha+1)\alpha - \alpha^2 \right] = \gamma^2\alpha$$

4. Maximum Likelihood Estimators and Equations

Although trying to fit a gamma distribution $\boldsymbol{\mathcal{G}}(\alpha, \gamma)$ to our sample sounds like a great idea, it so happens that we have no clue about the parameters to feed our model with (in fact, there is no way for us to find the true value of α and γ). The best solution so far is to find estimators $\hat{\alpha}$, $\hat{\gamma}$ (**that will always be positive !** this will be important for later computations) of our true parameters and this is where the **MLE** method comes in handy. However, before moving on to computations, let us start by introducing a few notations here. First, the definition of the likelihood as well as the log-likelihood (please mind that log denotes the natural logarithm).

$$\boldsymbol{\mathcal{L}}(\alpha,\gamma)=\prod_{i=1}^{n} f_{\boldsymbol{X}}(x_i;\alpha,\gamma) \qquad \boldsymbol{\mathcal{L}\mathcal{L}}(\alpha,\gamma)=\sum_{i=1}^{n}\log\Big(f_{\boldsymbol{X}}(x_i;\alpha,\gamma)\Big)$$

Finally, we denote by $\overline{\boldsymbol{X}} \overset{\textbf{set}}{=\!=} \frac{1}{n}\sum_{i=1}^{n} x_i$ the empirical mean of our sample and $\overline{\log(\boldsymbol{X})}=\frac{1}{n}\sum_{i=1}^{n}\log x_i$. Now that all is settled down, let us find the expression of our maximum likelihood estimators $(\hat{\alpha},\hat{\gamma}) \overset{\textbf{set}}{=\!=} \mathbf{argmax}_{(\alpha,\gamma)\in\boldsymbol{\Theta}}\boldsymbol{\mathcal{L}}(\alpha,\gamma)$. Our goal is obviously equivalent to finding a solution to the following problem:

$$\boxed{\textbf{(P) : Find } (\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\gamma}}) \textbf{ satisfying the Maximum Likelihood Equations, } \boldsymbol{i.e.} \quad \boldsymbol{\nabla\mathcal{L}\mathcal{L}(\alpha,\gamma)}\Big|_{(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\gamma}})}\boldsymbol{=0}}$$

Let's go!

$$\boldsymbol{\mathcal{L}}(\alpha,\gamma)=\prod_{i=1}^{n}\frac{1}{\boldsymbol{\Gamma}(\alpha)\gamma^{\alpha}}(x_i)^{\alpha-1}e^{-x_i/\gamma}=\boldsymbol{\Gamma}(\alpha)^{-n}\gamma^{-n\alpha}\exp\left(-\frac{1}{\gamma}\sum_{i=1}^{n}x_i\right)\left(\prod_{i=1}^{n}x_i\right)^{\alpha-1}$$

We can thus compute the log-likelihood, which yields

$$\boldsymbol{\mathcal{L}\mathcal{L}}(\alpha,\gamma)=-n\log\Big(\boldsymbol{\Gamma}(\alpha)\Big)-n\alpha\log\gamma-\frac{n}{\gamma}\overline{\boldsymbol{X}}+n\big(\alpha-1\big)\overline{\log(\boldsymbol{X})}$$

The log-likelihood is unconditionally differentiable on $\boldsymbol{\Theta}$ with respect to both variables so we can compute the partial derivatives along the two axes

$$\boldsymbol{\nabla\mathcal{L}\mathcal{L}}(\hat{\alpha},\hat{\gamma})=0 \iff \begin{cases}-n\frac{\boldsymbol{\Gamma}'(\hat{\alpha})}{\boldsymbol{\Gamma}(\hat{\alpha})}-n\log\hat{\gamma}+n\overline{\log(\boldsymbol{X})}=0\\ -\frac{n\hat{\alpha}}{\hat{\gamma}}+\frac{n\overline{\boldsymbol{X}}}{\hat{\gamma}^2}=0\end{cases}$$

Let us denote by $\boldsymbol{\psi}$ the digamma function $z\mapsto\boldsymbol{\Gamma}'(z)/\boldsymbol{\Gamma}(z)$. Our maximum likelihood equations are equivalent to the following non-linear system of two equations

$$\begin{cases}\boldsymbol{\psi}(\hat{\alpha})+\log\hat{\gamma}-\overline{\log(\boldsymbol{X})}=0\\ \hat{\alpha}\hat{\gamma}-\overline{\boldsymbol{X}}=0\end{cases} \iff \begin{cases}\boldsymbol{\psi}(\hat{\alpha})-\log(\hat{\alpha})=\overline{\log(\boldsymbol{X})}-\log(\overline{\boldsymbol{X}})\\ \hat{\gamma}=\overline{\boldsymbol{X}}/\hat{\alpha}\end{cases} \tag{1}$$

4.1. **Solution to Maximum Likelihood equations.** Let us now prove that system (1) admits a unique solution by first showing this sublemmas

Lemma 1. *The mapping* $\boldsymbol{\varphi}:\begin{cases}\mathbb{R}_+^*\longrightarrow\mathbb{R}_-^*\\ z\longmapsto\boldsymbol{\psi}(z)-\log(z)\end{cases}$ *defines a bijection.*

Proof. First, we show that $\boldsymbol{\varphi}$ is increasing. $\boldsymbol{\varphi}$ is differentiable on its domain, and its derivative $\boldsymbol{\varphi}'$ is equal to $\boldsymbol{\varphi}'(z)=\boldsymbol{\psi}'(z)-1/z=\sum_{k=0}^{+\infty}\frac{1}{(k+z)^2}-\frac{1}{z}$. Let us then define the integer part function as $[\cdot]:z\mapsto\sum_{k\in\mathbb{Z}}k\cdot\mathbf{1}_{[k,k+1[}(z)$ and notice that the following inequality holds for $\lambda-$almost every $z\in\mathbb{R}_+^*$:

$$\frac{1}{(k+z)^2} = \frac{1}{([t]+z)^2} > \frac{1}{(t+z)^2}$$

By taking the integral over $\mathbb{R}_+^*$ we obtain

$$\int_{\mathbb{R}_+^*} \frac{\mathrm{d}t}{([t]+z)^2} > \int_{\mathbb{R}_+^*} \frac{\mathrm{d}t}{(t+z)^2} \quad \Longleftrightarrow \quad \sum_{k\in\mathbb{N}} \frac{1}{(k+z)^2} \int_{[k,k+1]} \mathrm{d}t > \int_{[z,+\infty[} \frac{\mathrm{d}y}{y^2}$$

which leads to $\boldsymbol{\psi}'(z) > 1/z$ and thus $\boldsymbol{\varphi}'(z) = \boldsymbol{\psi}'(z) - \frac{1}{z} > \frac{1}{z} - \frac{1}{z} = 0$. Therefore $\boldsymbol{\varphi}$ is an increasing function. Additionally we have that $\lim_{z\to 0} \boldsymbol{\varphi}(z) = \lim_{z\to 0}\big[\psi(z) - \log(z)\big]$ which we can rewrite

$$\lim_{z\to 0}\left[-\boldsymbol{C} + \sum_{k\in\mathbb{N}}\left(\frac{1}{k+1} - \frac{1}{k+z}\right) - \log(z)\right] = -\boldsymbol{C} + \lim_{z\to 0}\left[\sum_{k\in\mathbb{N}} \frac{z-1}{(k+1)(k+z)} - \log(z)\right] = -\boldsymbol{C} + \lim_{z\to 0}\left[1 - \frac{1}{z} + (z-1)\boldsymbol{A}(z) - \log(z)\right]$$

where $\boldsymbol{A}(z) \overset{\textbf{set}}{=\!=} \sum_{k\geq 1}(k+1)^{-1}(k+z)^{-1}$ is a positive quantity, meaning that as soon as z gets smaller than 1, the quantity $(z-1)\boldsymbol{A}(z)$ become negative. This enables us to write $\lim_{z\to 0}\boldsymbol{\varphi}(z) < 1 - 1/z - \log(z)$ that clearly goes to $-\infty$ as z gets closer to 0. Finally, when z goes to $+\infty$, we can look at the limit on integer quantities by growth comparison and thus

$$\lim_{z\to+\infty} \boldsymbol{\varphi}(z) = \lim_{n\to+\infty}\left[-\boldsymbol{C} + \sum_{k\in\mathbb{N}}\left(\frac{1}{k+1} - \frac{1}{k+n}\right) - \log(n)\right] = -\boldsymbol{C} + \lim_{n\to+\infty}\left[\sum_{k=0}^{n-1}\frac{1}{k+1} - \log(n)\right]$$

We instantly recognize the expression $\boldsymbol{H}(n-1) - \log(n)$ (harmonic series minus the logarithm) under the limit operator which, with n growing, goes to $\boldsymbol{C}$ (the **Euler-Mascheroni** constant). Hence by summation, $\lim_{z\to+\infty}\boldsymbol{\varphi}(z) = 0$. To sum up, $\boldsymbol{\varphi}$ is an increasing functions whose limits at 0 and $+\infty$ are $-\infty$ and 0 respectively, therefore $\boldsymbol{\varphi}$ is a bijection as wanted.

QED

Still, we have to be careful about an extreme case that we should consider in order to be completely exhaustive. In the case where only one value, say x_0 is taken by each $\boldsymbol{X}_i$, we will have $\log(\overline{\boldsymbol{X}}) = \overline{\log(\boldsymbol{X})}$, in which case our inversion would yield $\hat{\alpha} = +\infty$. Recall that in that case, we have $\hat{\gamma} = \frac{x_0}{\hat{\alpha}}$. Thus, one can show the additional lemma

Lemma 2. *if a random variable Y_α follows a Gamma distribution $\boldsymbol{\mathcal{G}}\big(\alpha, \frac{x_0}{\alpha}\big)$, then Y_α converges in probability to a variable $\boldsymbol{Y}$ with **Dirac** distribution δ_{x_0}.*

Proof. We need to show that for any $\varepsilon > 0$, the quantity $\mathbb{P}\Big[|Y_\alpha - x_0| \geq \varepsilon\Big] \longrightarrow 0$ as α goes to infinity.

$$\mathbb{P}\Big[|Y_\alpha - x_0| \geq \varepsilon\Big] = \int_0^{x_0-\varepsilon} \frac{1}{\Gamma(\alpha)(\frac{x_0}{\alpha})^\alpha} t^{\alpha-1} e^{-\alpha t/x_0}\mathrm{d}t + \int_{x_0+\varepsilon}^{\infty} \frac{1}{\Gamma(\alpha)(\frac{x_0}{\alpha})^\alpha} t^{\alpha-1} e^{-\alpha t/x_0}\mathrm{d}t$$

Thus, we work to rewrite the integrand in a friendlier form :

$$\begin{aligned}
\frac{1}{\boldsymbol{\Gamma}(\alpha)\left(\frac{x_0}{\alpha}\right)^\alpha} t^{\alpha-1} e^{-\alpha t/x_0} &= \frac{1}{x_0}\cdot\frac{1}{\boldsymbol{\Gamma}(\alpha)\alpha^{-\alpha}}\left(\frac{t}{x_0}\right)^{\alpha-1} e^{-\alpha t/x_0} \\
&= \frac{1}{x_0} e^{-t/x_0} \frac{\alpha^\alpha}{\boldsymbol{\Gamma}(\alpha)}\left(\frac{t}{x_0} e^{-t/x_0}\right)^{\alpha-1}
\end{aligned}$$

A quick study of the $s \mapsto se^{-s}$ function reveals that it has a strict maximum of e^{-1} at $s = 1$, being strictly increasing before and strictly decreasing after, meaning that for every $\eta > 0$, there exists a $q > e$ such that $se^{-s} \leq q^{-1}$ for $|s - 1| \geq \eta$. Here, we set $\eta = \frac{\varepsilon}{x_0}$ and $s = t/x_0$, yielding :

$$\frac{1}{\mathbf{\Gamma}(\alpha)\left(\frac{x_0}{\alpha}\right)^{\alpha}} t^{\alpha-1} e^{-\alpha t/x_0} \leq \left[\frac{q}{x_0} e^{-t/x_0}\right] \cdot \left[\frac{\alpha^{\alpha}}{\mathbf{\Gamma}(\alpha)} q^{-\alpha}\right]$$

Notice that the first term is a function that is integrable over $\mathbb{R}_+$ that *does not depend on* α, while the second one is a constant that does not depend on t. We conclude by using Stirling's formula :

$$\mathbf{\Gamma}(\alpha + 1) \sim \sqrt{2\pi\alpha}\left(\frac{\alpha}{e}\right)^{\alpha}$$

Therefore :

$$\frac{\alpha^{\alpha}}{\mathbf{\Gamma}(\alpha)} q^{-\alpha} \sim (\alpha + 1)\frac{\alpha^{\alpha}}{\alpha^{\alpha}\sqrt{2\pi\alpha}}\left(\frac{e}{q}\right)^{\alpha} \sim \frac{\alpha + 1}{\sqrt{2\pi\alpha}}\left(\frac{e}{q}\right)^{\alpha}$$

Since $e/q < 1$ and q does not depend on α, this term vanishes when $\alpha \to \infty$. It is, in particular, bounded, meaning that there exists a constant $\boldsymbol{A}(q)$ depending only on q, such that $\frac{\alpha^{\alpha}}{\Gamma(\alpha)} q^{-\alpha} \leq \boldsymbol{A}(q)$. From this, we deduce that $t \mapsto \frac{1}{\mathbf{\Gamma}(\alpha)(\frac{x_0}{\alpha})^{\alpha}} t^{\alpha-1} e^{-\alpha t/x_0}$ converges pointwise to zero for $|t - x_0| \geq \varepsilon$, and is bounded by $\boldsymbol{B}(x_0, \varepsilon)e^{-t/x_0}$ with $\boldsymbol{B}(x_0, \varepsilon) = \frac{q\boldsymbol{A}(q)}{x_0}$

The **Dominated Convergence Theorem** implies that, as claimed, $\mathbb{P}\left[|Y_{\alpha} - x_0| \geq \varepsilon\right] \longrightarrow 0$ as α gets larger and larger.

QED

The previous lemma explains the seemingly bizarre "$\hat{\alpha} = \infty$" result obtained if all the randomly drawn values are the same : in that case, the model suggests that the most likely distribution is a dirac δ distribution, which, both plot-wise and theoretically, makes perfect sense. Now, using the concavity of log, and **Jensen**'s inequality, we obtain $\overline{\log(\boldsymbol{X})} - \log(\overline{\boldsymbol{X}}) < 0$. It is then only left to conclude that, given any sampling of X, there exists a unique $\hat{\alpha}$ satisfying $\hat{\alpha} = \boldsymbol{\varphi}^{-1}(\overline{\log(\boldsymbol{X})} - \log(\overline{X}_n))$ which at the same time, defines a unique $\hat{\gamma} = \overline{\boldsymbol{X}}/\hat{\alpha}$. Those $\hat{\alpha}$ and $\hat{\gamma}$ are our unique solution to the Maximum Likelihood Equations (1)

4.2. **Double check of the uniqueness of the Maximum Likelihood Estimator.** Even though we are sure that our Maximum Likelihood Equations admits a unique **MLE**, a double check won't be of any harm. A sufficient condition to prove uniqueness, is to show that the hessian matrix $\boldsymbol{\mathcal{H}}_{\boldsymbol{\mathcal{LL}}}$ is negative definite where the gradient $\boldsymbol{\nabla}\boldsymbol{\mathcal{LL}}$ vanishes. To make sure this property is satisfied, let us compute the hessian matrix of $\boldsymbol{\mathcal{LL}}$.

$$\boldsymbol{\mathcal{H}}_{\boldsymbol{\mathcal{LL}}}(\alpha, \gamma) \overset{\textbf{set}}{=\!=} \begin{bmatrix} \partial_{\alpha}^2 \boldsymbol{\mathcal{LL}}(\alpha, \gamma) & \partial_{\alpha}\partial_{\gamma}\boldsymbol{\mathcal{LL}}(\alpha, \gamma) \\ \partial_{\gamma}\partial_{\alpha}\boldsymbol{\mathcal{LL}}(\alpha, \gamma) & \partial_{\gamma}^2\boldsymbol{\mathcal{LL}}(\alpha, \gamma) \end{bmatrix} = \begin{bmatrix} -n\boldsymbol{\psi}'(\alpha) & -n/\gamma \\ -n/\gamma & (n\alpha\gamma - 2n\overline{\boldsymbol{X}})/\gamma^3 \end{bmatrix}$$

Now assume $(\hat{\alpha}, \hat{\gamma})$ is a solution to the likelihood equations. By definition, they obey to equations (1), in particular, $\hat{\alpha} > 0$, $\hat{\gamma} > 0$ and $\hat{\alpha}\hat{\gamma} = \overline{\boldsymbol{X}}$. In such a case, we are guaranteed that each entry of $\boldsymbol{\mathcal{H}}_{\boldsymbol{\mathcal{LL}}}$ is negative, indeed $-n\boldsymbol{\psi}'(\hat{\alpha}) = -n\sum_{k\in\mathbb{N}}(k + \hat{\alpha})^{-2} < 0$, $-\frac{n}{\hat{\gamma}} < 0$ and finally notice that

$$\frac{n\hat{\alpha}\hat{\gamma} - 2n\overline{\boldsymbol{X}}}{\gamma^3} = \frac{n\overline{\boldsymbol{X}} - 2n\overline{\boldsymbol{X}}}{\gamma^3} = -\frac{n\overline{\boldsymbol{X}}}{\gamma^3} < 0$$

Thus, $\boldsymbol{\mathcal{H}}_{\boldsymbol{\mathcal{LL}}}(\hat{\alpha}, \hat{\gamma})$ is negative definite and the **MLE** $(\hat{\alpha}, \hat{\gamma})$ is unique.

5. Numerical Methods for Likelihood Estimators

It happens that our maximum likelihood estimators does not admit a simple closed form. However it is still possible to get a great approximation of the true values using numerical methods. We will use two here, but notice there exists many others (such as **Regula-Falsi**, the Fixed Point method, dichotomy, ...). Here, notice we have to estimate two parameters, meaning that we shall use methods in higher dimension. However, one can remark that the parameter $\hat{\gamma}$ is uniquely determined by $\hat{\alpha}$ through the relation $\hat{\gamma} = \overline{\boldsymbol{X}}/\hat{\alpha}$. Therefore, it is enough to apply a numerical method of our choice to α which will greatly reduce the complexity of computations. Let us use in the following **Newton** and **Fisher** methods.

5.1. **Newton-Raphson Equations.**

From system (1), we seek for $\hat{\alpha}$ the root of the equation

$$\boldsymbol{\varphi}(\alpha) \stackrel{\textbf{set}}{=\!=} \boldsymbol{\psi}(\alpha) - \log(\alpha) + \log(\overline{\boldsymbol{X}}) - \overline{\log(\boldsymbol{X})} = 0$$

Previous computations shows that the derivative $\boldsymbol{\varphi}'$ of $\boldsymbol{\varphi}$ does not vanish on its domain, therefore the quantity $1/\boldsymbol{\varphi}'$ always makes sense. Thus applying **Newton**'s method to $\boldsymbol{\varphi}$ yields the following iterative formula.

$$\begin{cases} \alpha^{(k+1)} & = & \alpha^{(k)} - \frac{\boldsymbol{\psi}\left(\alpha^{(k)}\right) - \log\left(\alpha^{(k)}\right) + \log(\overline{\boldsymbol{X}}) - \overline{\log(\boldsymbol{X})}}{\boldsymbol{\psi}'\left(\alpha^{(k)}\right) - 1/\alpha}, \quad k \geq 0 \\ \alpha^{(0)} & = & \alpha_0 \in \mathbb{R}^+ \end{cases}$$

Rather than computing each iteration by hand, we will let the computer do the work using the piece of code located in Appendix [A]. Please keep in mind that the method will not always converge. We thus need to carefully choose our starting point that will be sufficiently close to the actual value. The usual way to proceed when dealing with functions having complex expressions is to use approximations such as first order **Taylor** expansion of $\boldsymbol{\varphi}$ at $z = 1$ (a plot of the curve $(z, \boldsymbol{\varphi}(z))$ justifies our choice).

$$\begin{aligned} \boldsymbol{\varphi}(z) &\approx \boldsymbol{\varphi}(1) + \frac{\boldsymbol{\varphi}'(1)}{2}(z-1) + \log(\overline{\boldsymbol{X}}) - \overline{\log(\boldsymbol{X})} \\ &\approx -\boldsymbol{C} + \frac{(\pi^2 - 6)}{6}(z-1) + \log(\overline{\boldsymbol{X}}) - \overline{\log(\boldsymbol{X})} \end{aligned}$$

Which, if we plug data from the sample and solve for z yields the choice $\alpha^{(0)} \approx 1.45$. When running the script, we obtain the value of $\hat{\alpha} = 1.915043$ with the method converging with a satisfactory precision in only four iterations. Plugging this information into the equation satisfied by gamma yields $\hat{\gamma} = 135.1573$ that are the two parameters obtained with the package `libdistrplus` which is quite reassuring. Additional fancy visuals can be obtained, using the package `animation` in `R`, that allows us to plot the values at each iteration of the method.

5.2. **Fisher Equations.**

As mentioned at the beginning of previous section, **Newton**'s method is not the only one we can use to compute our maximum likelihood estimators, for the sake of illustration, we will move on to another method using the information matrix whose definition is to be recalled in the multidimensional case for $\boldsymbol{\theta} = (\alpha, \gamma)$.

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) = -\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^T}\log\prod_{i=1}^{n} f(X_i \mid \boldsymbol{\theta})\right]$$

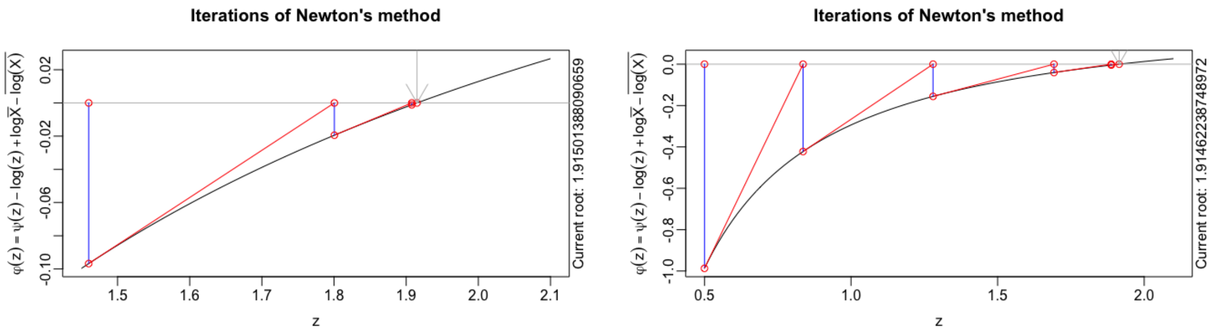

FIGURE 7. Two runs of **Newton**'s method with initial values $\alpha_1^{(0)} = 1.45$, and $\alpha_2^{(0)} = 0.5$

In our case, we get the following information matrix

$$
\begin{aligned}
\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) &= -\mathbb{E}_{\boldsymbol{\theta}}\left[\boldsymbol{\mathcal{H}}_{\boldsymbol{\mathcal{LL}}}(\boldsymbol{\theta})\right] \\
&= n \begin{bmatrix} \mathbb{E}[\boldsymbol{\psi}'(\alpha)] & \mathbb{E}[1/\gamma] \\ \mathbb{E}[1/\gamma] & \mathbb{E}[2\overline{\boldsymbol{X}}/\gamma^3 - \alpha/\gamma^2] \end{bmatrix} \\
&= n \begin{bmatrix} \psi'(\alpha) & 1/\gamma \\ 1/\gamma & \mathbb{E}[X_1]/\gamma^3 - \alpha/\gamma^2 \end{bmatrix} \\
&= n \begin{bmatrix} \boldsymbol{\psi}'(\alpha) & 1/\gamma \\ 1/\gamma & 2\alpha\gamma/\gamma^3 - \alpha/\gamma^2 \end{bmatrix} \\
&= n \begin{bmatrix} \boldsymbol{\psi}'(\alpha) & 1/\gamma \\ 1/\gamma & \alpha/\gamma^2 \end{bmatrix}
\end{aligned}
$$

Fisher equations are defined precisely as the **Newton-Raphson** equations up to one difference: instead of using the jacobian matrix (or derivative in the unidimensional case), we plug $-\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\theta})$ ($\boldsymbol{\mathcal{I}}$ is of course invertible), which results in the following iterative method

$$
\begin{cases}
\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\theta}) \cdot \boldsymbol{\nabla}\boldsymbol{\mathcal{LL}}\left(\boldsymbol{\theta}^{(k)} \mid \boldsymbol{X}\right) \\
\boldsymbol{\theta}^{(0)} = \left(\alpha^{(0)}, \gamma^{(0)}\right)
\end{cases}
$$

We can rewrite the iterative formula in its fully expanded form to obtain a more explicit equation

$$
\begin{bmatrix} \alpha^{(k+1)} \\ \gamma^{(k+1)} \end{bmatrix} = \begin{bmatrix} \alpha^{(k)} \\ \gamma^{(k)} \end{bmatrix} - \frac{1}{\alpha^{(k)}\boldsymbol{\psi}'\left(\alpha^{(k)}\right) - 1} \begin{bmatrix} \alpha^{(k)} & -\gamma^{(k)} \\ -\gamma^{(k)} & \left(\gamma^{(k)}\right)^2 \psi'\left(\alpha^{(k)}\right) \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\psi}\left(\alpha^{(k)}\right) + \log\left(\gamma^{(k)}\right) - \overline{\log(\boldsymbol{X})} \\ \alpha^{(k)}/\gamma^{(k)} - \overline{\boldsymbol{X}}/\left(\gamma^{(k)}\right)^2 \end{bmatrix}
$$

This expression being quite awful to evaluate manually, we will once again compute each step numerically using the piece of code in appendix [B]. Without surprise, the code returns expected parameters that are identical to those found with **Newton** method, *i.e.* $\hat{\alpha} = 1.915, \hat{\gamma} = 135.158$

6. Asymptotic distribution analysis

Although we already had an idea of the best parameters thanks to **Newton**'s method, computing the **Fisher** information matrix $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})$ was not in vain. Indeed, we will make use of this matrix in order to deduce the asymptotic distribution of our parameters. But first, we need to check that they indeed have an asymptotic distribution. For that matter, we need to check some things first. We already have that $\mathfrak{M}$ is a regular parametric model, thus it is enough to check the following hypothesis

6.1. **Satisfaction of H1'.** This is an easy one. Let $\boldsymbol{\theta}_0 = (\alpha_0, \gamma_0)$ be our true parameter and notice that $\boldsymbol{\theta}_0 \in \mathring{\boldsymbol{\Theta}} = \boldsymbol{\Theta}$ (except in the case where all observations $X_i \in \boldsymbol{X}$ are identical, but the latter is to be dismissed from our study of asymptotic behavior).

6.2. **Satisfaction of H2'.** It is sufficient to remark that our parameter space $\boldsymbol{\Theta}$ is an open set and that $\boldsymbol{\theta} \mapsto f_{\boldsymbol{X}}(x, \boldsymbol{\theta})$ is clearly twice differentiable as a product and composition of differentiable functions on $\mathbf{Dom}(f_{\boldsymbol{X}})$.

6.3. **Satisfaction of H3'.** For any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, recall the following results

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) = n \begin{bmatrix} \boldsymbol{\psi}'(\alpha) & 1/\gamma \\ 1/\gamma & \alpha/\gamma^2 \end{bmatrix}, \qquad \boldsymbol{\nabla_\theta} \log f\left(\boldsymbol{X}, \boldsymbol{\theta}\right) = \begin{bmatrix} -\boldsymbol{\psi}(\alpha) - \log(\gamma) + \log(\boldsymbol{X}) \\ -\alpha/\gamma + \boldsymbol{X}/\gamma^2 \end{bmatrix}$$

Firstly, it holds true that, $\det\left[\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)\right] = n^2\left[\alpha_0\boldsymbol{\psi}'(\alpha_0) - 1\right]/\gamma_0^2$ with $\gamma_0 \neq 0$, furthermore,

$$\alpha_0\psi'(\alpha_0) - 1 = \alpha_0 \sum_{k=0}^{+\infty} \frac{1}{(k+\alpha_0)^2} - 1 = \alpha_0 \sum_{k=0}^{+\infty} \int_{[k,k+1]} \frac{\mathrm{d}t}{([t]+\alpha_0)^2} - 1 > \int_{[\alpha_0,+\infty]} \frac{\alpha_0}{\tau^2}\mathrm{d}\tau - 1 = \left[\frac{\alpha_0}{\alpha_0}\right] - 1 = 0$$

which proves the positivity of the quantity $\det\left[\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)\right]$, the **Fisher** matrix is thus non-singular. Additionally

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left\|\boldsymbol{\nabla_\theta} \log f(\boldsymbol{X}, \boldsymbol{\theta}_0)\right\| = \mathbb{E}\left[\left[\boldsymbol{\nabla_\theta} \log f(\boldsymbol{X}, \boldsymbol{\theta}_0)\right]^{\boldsymbol{T}} \left[\boldsymbol{\nabla_\theta} \log f(\boldsymbol{X}, \boldsymbol{\theta}_0)\right]\right]$$

for the sake of space saving, let us denote by $\boldsymbol{Y}$ the quantity $\boldsymbol{\nabla_\theta} \log f(\boldsymbol{X}, \boldsymbol{\theta}_0)$. Then it holds true that

$$\boldsymbol{Y^T Y} = \left(\underbrace{-\boldsymbol{\psi}(\alpha_0) - \log(\gamma_0)}_{:=c_1} + \log(\boldsymbol{X})\right)^2 + \left(\underbrace{-\alpha_0/\gamma_0}_{:=c_2} + \boldsymbol{X}/\gamma_0^2\right)^2$$

expanding the two squares and taking the expectation gives us

$$\mathbb{E}_{\boldsymbol{\theta}_0}[\boldsymbol{Y^T}Y] = c_1^2 + 2c_1\mathbb{E}[\log \boldsymbol{X}] + \mathbb{E}[\log^2 \boldsymbol{X}] + c_2^2 + \frac{2c_2}{\gamma_0^2}\mathbb{E}[\boldsymbol{X}] + \frac{1}{\gamma^4}\mathbb{E}[\boldsymbol{X}^2]$$

Finally, it follows that $\mathbb{E}_{\boldsymbol{\theta}_0}\left\|\boldsymbol{\nabla_\theta} \log f_{\boldsymbol{X}}(x, \boldsymbol{\theta}_0)\right\| < \infty$.

6.4. **Satisfaction of H4'.** The first condition of **H4'** is proven exactly the same way as in **H3'** and we will thus skip the computation. The other two conditions are not verified - but they are mainly used to prove that the matrix $\nabla^2 \bar{l}_n(\boldsymbol{\theta}_0)^{-1} \to -\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$, which we can check manually.

With all four hypotheses satisfied by $\mathfrak{M}$, we can finally conclude that our maximum likelihood estimator $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\gamma})$ will be asymptotically normally distributed according to the following

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \longrightarrow \mathcal{N}\left(0,\ \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)^{-1}\right) \approx \mathcal{N}\left(0,\ \boldsymbol{\mathcal{I}}(\hat{\boldsymbol{\theta}})^{-1}\right)$$

Since $\hat{\gamma}$ is quite large in our dataset, we can assume that the fisher information matrix has non-diagonal coefficients $1/\hat{\gamma} \approx 0$, thus

$$\mathbf{Var}(\hat{\alpha}) \approx \frac{\hat{\alpha}}{n(\hat{\alpha}\boldsymbol{\psi}'(\hat{\alpha}) - 1)} = 0.046$$
$$\mathbf{Var}(\hat{\gamma}) \approx \frac{\hat{\gamma}^2 \psi'(\hat{\alpha})}{n(\hat{\alpha}\boldsymbol{\psi}'(\hat{\alpha}) - 1)} = 298.35$$

Where we can observe that there is a huge variance on $\hat{\gamma}$, which is pretty bad regarding to the actual results obtained *via* the computer but this is not too worrying because it does not allow $\hat{\gamma}$ to be negative. In any case, this study concludes our asymptotical analysis section.

7. Evolution of the Mean-Squared Error (MSE) for Maximum-Likelihood Estimators

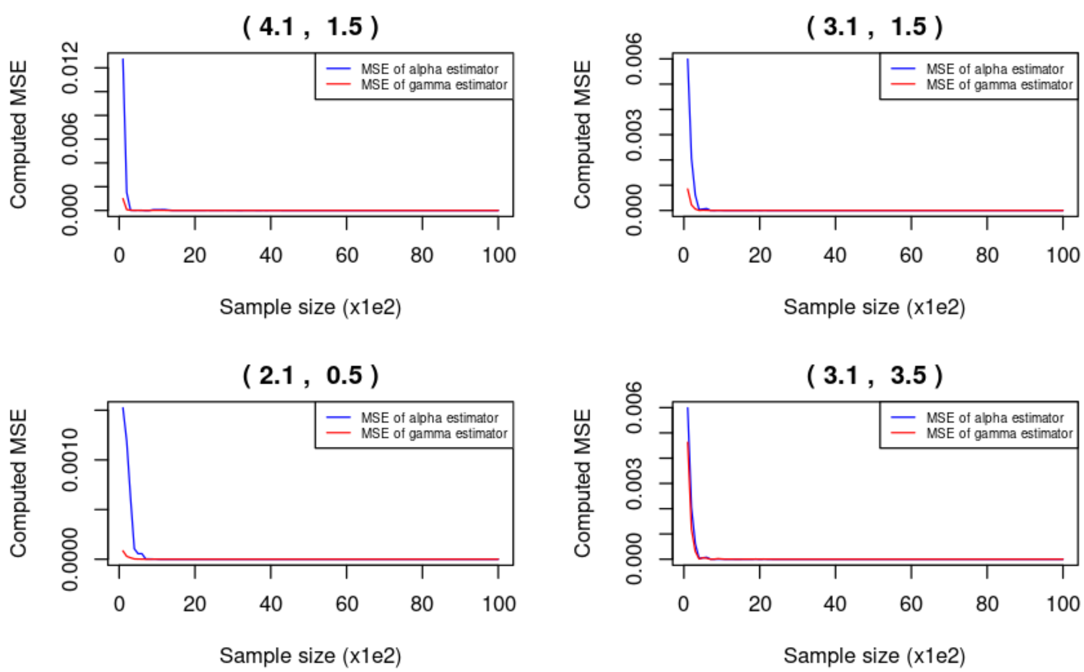

Figure 8. MSE of estimators $\hat{\alpha}$ and $\hat{\gamma}$ as a function of sample size

The MSE for the estimators of the parameters of the gamma distribution decreases exponentially when the size of the samples are increased. Such evolution is expected since sample size is generally affected by or associated with the latter's standard error, statistical power and confidence level. Starting from a sample size of 10^2, the MSE decreases drastically as we approach a sample size of 10^5 and becomes asymptotic. In

Fig. 8, from left to right, we take $(\hat{\alpha}, \hat{\gamma}) = (4.1, 1.5)$, $(3.1, 1.5)$, $(2.1, 0.5)$ and $(3.1, 3.5)$ (randomly generated) and visualise the MSE's evolution within a range of 10^2 and 10^4.

8. Conclusion

In this experimentation, we first modeled the empirical information based on different statistical distributions at our disposal. Several statistical methods were employed to narrow down different distribution options. We conclude that out of all the distributions, the Gamma distribution works the best. Although it was not perfectly compatible, we discussed its statistical implications. We find that the arguably best distribution for this analysis would be the **Box-Cox exponential distribution**, which is beyond the scope of this experimentation. Once the validity of the Gamma distribution was established, we performed statistical computations with that presumption to mathematically argue and calculate the statistical compatibility of the model chosen, based on the computations of Maximum-Likelihood estimation.

Appendix A. R code for Newton Raphson numerical method

```
setwd(dir="~/Desktop") # data file is located on the desktop
data <- read.table("gpigs.txt", sep=";", header=T) # import data

b <- data$censored==0 # find uncensored data
uncensored <- data[b, ] # only keep uncensored data

X <- uncensored$lifetime # store our sample in X

X.bar <- mean(X) # compute the mean of the sample
X.log <- log(X) # take the log of the sample
X.log.bar <- mean(X.log) # take the mean of log(sample)

a <- log(X.bar) - X.log.bar # the constant in function phi

phi <- function(z, a) return(digamma(z) - log(z) + a) # phi function...

phi.diff <- function(z) return(trigamma(z) - 1/z) # ...and its derivative

## Setup Newton-Raphson method
iter <- 10 # how many times should we iterate through the method
x.old <- 1  # initial value
comp.alpha <- numeric(length = iter); comp.alpha[1] <- x.old # vector of estimated alpha

## Apply Newton-Raphson method
for(i in 2:iter) {
  x.new <- x.old - phi(x.old, a) / phi.diff(x.old)
  x.old <- x.new
  comp.alpha[i] <- x.new
}


comp.gamma <- X.bar / comp.alpha # vector of estimated gamma

comp.alpha[iter] # print alpha
comp.gamma[iter] # print gamma
```

Appendix B. R code for Fisher numerical Method

```r
setwd("~/Documents/cmb/semester-2/advanced-stats/project")
data.gpigs <- read.table("surv.gpigs.txt", header = TRUE, sep = ";")
data.gpigs$regime <- factor(data.gpigs$regime)

data.gpigs_cens <- data.gpigs[which(data.gpigs$censored == 0), ] #retrieve data
X <- data.gpigs_cens$lifetime # sample
n <- length(X)
Xbar <- mean(X)
logXbar <- mean(log(X))

# function for scoring
S <- function(a, g) {
  m <- matrix(data = c(digamma(a) + log(g) - logXbar, a/g - Xbar/g^2), nrow = 2, ncol = 1)
  return(-n * m)
}

I <- function(a, g) {
  m <- matrix(data = c(c(trigamma(a), 1/g), c(1/g, a/g^2)), nrow = 2, ncol = 2)
  return(n * m)
}

# simulations

a_old <- 1.5
g_old <- 130
a_g_old <- matrix(data = c(a_old, g_old), nrow = 2, ncol = 1)

for(i in 1:10) {
  new_a_g <- a_g_old + solve(I(a_old, g_old)) %*% S(a_old, g_old)
  a_old <- new_a_g[1,]
  g_old <- new_a_g[2,]
  a_g_old <- matrix(data = c(a_old, g_old), nrow = 2, ncol = 1)
}

new_a_g
```

Appendix C. R code for Monte Carlo experiments to study the evolution of MSE of Maximum-Likelihood estimators

```r
N <- seq(1e2, 1e4, by = 1e2) # sample sizes
s <- 4 # parameter options
palpha <- sample(1.1:5, s, replace = TRUE) # alpha samples
pgamma <- sample(0.5:3.5, s, replace = TRUE) # gamma samples
true.alpha <- matrix(palpha, nrow = length(N), ncol = s,
                     byrow = TRUE) # combine all alpha samples
true.gamma <- matrix(pgamma, nrow = length(N), ncol = s,
                     byrow = TRUE) # combine all gamma samples
comp.alpha <- matrix(data = NA, ncol = s, nrow = length(N)) # matrix for storing computed alpha
comp.gamma <- matrix(data = NA, ncol = s, nrow = length(N)) # matrix for storing computed gamma
for(n in N) {
  for(i in 1:s) {
    set.seed(010)
    X <- rgamma(n, scale = pgamma[i], shape = palpha[i]) # generate a gamma sample of size n
    X.bar <- mean(X) # compute the mean of the sample
    X.log <- log(X) # take the log of the sample
    X.log.bar <- mean(X.log) # take the mean of log(sample)
    a <- log(X.bar) - X.log.bar
    phi <- function(z, a) return(digamma(z) - log(z) + a)
    phi.diff <- function(z) return(trigamma(z) - 1/z)
    ## Newton-Raphson method for estimation
    iters <- 15
    x.old <- mean(palpha)  # initial value
    vec.alpha <- numeric(length = iters); vec.alpha[1] <- x.old
    for(it in 2:iters) {
      x.new <- x.old - phi(x.old, a) / phi.diff(x.old)
      x.old <- x.new
      vec.alpha[it] <- x.new
    }
    vec.gamma <- X.bar / vec.alpha # vector of estimated gamma
    errors <- (vec.alpha - palpha[i]) ^ 2 + (vec.gamma - pgamma[i]) ^ 2 # computing the error
    comp.alpha[which(N == n), i] <- vec.alpha[length(vec.alpha)]
    comp.gamma[which(N == n), i] <- vec.gamma[length(vec.gamma)]
  }
}
comp.a.MSE <- ((true.alpha - comp.alpha) ^ 2) / N # compute MSE for alpha estimator
comp.g.MSE <- ((true.gamma - comp.gamma) ^ 2) / N # compute MSE for gamma estimator
# Plotting the evolution
par(mfcol = c(2, 2));
for(i in 1:s) {
  plot(comp.a.MSE[, i], type = "l", col = "blue",
       xlab = "Sample size (x1e2)", ylab = "Computed MSE",
       main = paste("(", palpha[i], ", ", pgamma[i], ")"))
  #text(1:s, comp.a.MSE[, i] + 4e-7, comp.a.MSE[, i], cex = 0.5, col = "blue")
  lines(comp.g.MSE[, i], type = "l", col = "red")
  legend("topright", legend = c("MSE of alpha estimator", "MSE of gamma estimator"),
         col = c("blue", "red"), lty = 1, cex = 0.6)
}
```

Appendix D. (Bonus). Other Maximum Likelihood Estimators

D.1. **Log-Normal fitting.** Recall the p.d.f. of the log-normal distribution

$$f(x;\mu,\sigma)=\frac{1}{x}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\left(\log x-\mu\right)^2/2\sigma^2\right]$$

One can compute the **MLE** as follows

$$\mathcal{L}(\mu,\sigma)=\prod_{i=1}^{n}f(x_i;\mu,\sigma)=(\sqrt{2\pi}\sigma)^{-n}\prod_{i=1}^{n}x_i^{-1}e^{-\frac{(\log x-\mu)^2}{2\sigma^2}}$$

$$\mathcal{LL}(\mu,\sigma)=-n\left(1+\frac{\mu}{\sigma^2}\right)\overline{\log(\boldsymbol{X})}+n\frac{\mu^2}{2\sigma^2}-\frac{n}{2\sigma^2}\overline{\log^2\boldsymbol{X}}-n\log(\sigma)-\frac{n}{2}\log(2\pi)$$

with our maximum likelihood estimators solving the equation

$$\nabla_{(\mu,\sigma)}\mathcal{LL}(\mu,\sigma)\Big|_{(\hat{\alpha},\hat{\gamma})}=0$$

Which, when we try to fit it using the `libdistr` package, gives the following fitting. Applying **Newton**'s method to our **MLE** equations yields the same parameters.

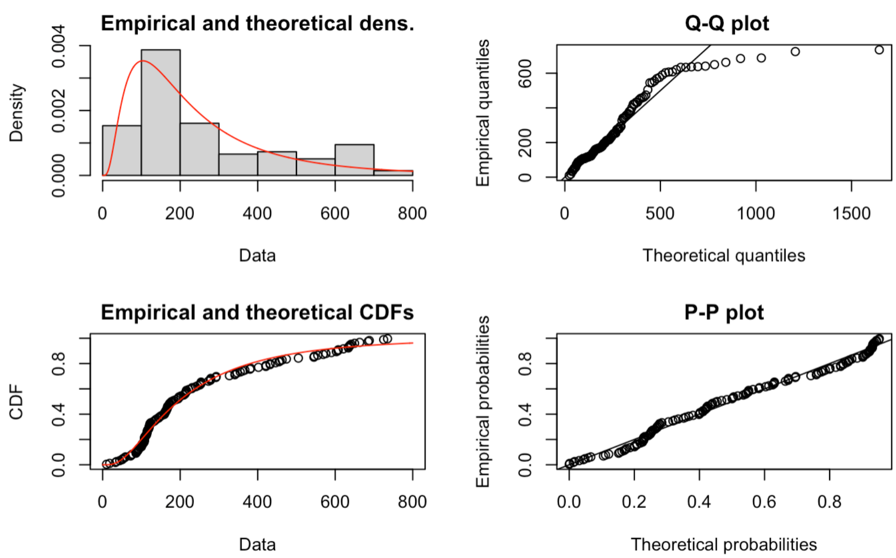

Figure 9. Fitting a log-normal distribution to our sample

D.2. **Log-Logistic fitting.** Recall the p.d.f. of the log-logistic distribution

$$f(x;\alpha,\beta) = \frac{\beta}{\alpha}\frac{(x/\alpha)^{\beta-1}}{[1+(x/\alpha)^{\beta}]^2}$$

we, similarly compute the MLE which we can compare the the fitting provided by `fitdistrplus` and remark that everything works nicely.

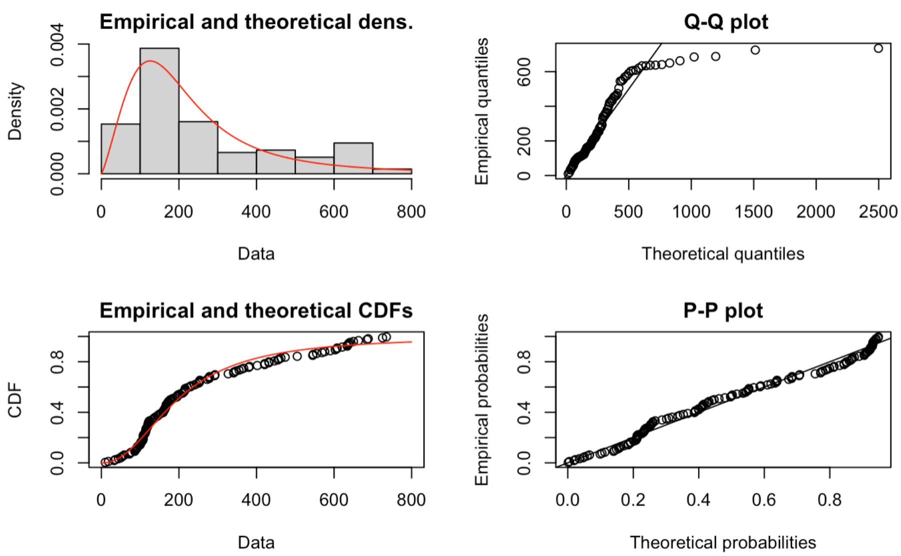

FIGURE 10. Fitting a log-logistic distribution to our sample

References

[1] Steve M Bajgier and Lalit K Aggarwal. "Powers of goodness-of-fit tests in detecting balanced mixed normal distributions". In: *Educational and Psychological Measurement* 51.2 (1991), pp. 253–269.

[2] Tor Bjerkedal et al. "Acquisition of Resistance in Guinea Pies infected with Different Doses of Virulent Tubercle Bacilli." In: *American Journal of Hygiene* 72.1 (1960), pp. 130–48.

[3] Alison C Cullen, H Christopher Frey, and Christopher H Frey. *Probabilistic techniques in exposure assessment: a handbook for dealing with variability and uncertainty in models and inputs*. Springer Science & Business Media, 1999.