LINMA1731 – Project 2019
Fish schools tracking

By Louis Navarre and Gilles Peiffer

Abstract

In this paper we solve the first part of the project for the class "Stochastic processes: Estimation and prediction" given during the Fall term of 2019. The average speed of each fish in a school of fish is approximated by a gamma-distributed random variable with a shape parameter k and a scale parameter s, and various methods for estimating this quantity are given; a numerical simulation is also included.

Contents

Part 1. Average speed estimation

1. Introduction

For the purpose of this project, we assume that the speed of each fish in a school at time i is a random variable V_i following a Gamma distribution, as suggested in [1]. This distribution is characterized by two parameters: a shape parameter $k > 0$ and a scale parameter $s > 0$. The parameters are the same for every fish and are time invariant. The aim of this first part is to identify these two parameters using empirical observations v_i.

2. A maximum likelihood estimator for a scalar parameter

Let v_i be i.i.d. realisations of a random variable following a Gamma distribution $\Gamma(k, s)$ (with $i = 1, \ldots, N$). We first assume that the shape parameter k is known.

We start by deriving the maximum likelihood estimator of $\vartheta := s$ based on N observations. Since the estimand ϑ is a deterministic quantity, we use Fisher estimation. In order to do this, let us restate the probability density function of $V_i \sim \Gamma(k, s)$:

$$f_{V_i}(v_i; k, s) = \frac{1}{\Gamma(k)s^k} v_i^{k-1} \mathrm{e}^{-\frac{v_i}{s}}, \quad i = 1, \ldots, N. \tag{2.1}$$

With this in mind, we can find that the likelihood $\mathcal{L}(v_1, \ldots, v_N; k, \vartheta)$ is given by

$$\mathcal{L}(v_1, \ldots, v_N; k, \vartheta) = \prod_{i=1}^{N} f_{V_i}(v_i; k, \vartheta) = \prod_{i=1}^{N} \frac{1}{\Gamma(k)\vartheta^k} v_i^{k-1} \mathrm{e}^{-\frac{v_i}{\vartheta}}. \tag{2.2}$$

In order to alleviate notation, we compute instead the log-likelihood, which is generally easier to work with[1]:

$$\begin{aligned} \ell(v_1, \ldots, v_N; k, \vartheta) &:= \ln \mathcal{L}(v_1, \ldots, v_N; k, \vartheta) && (2.3)\\ &= \sum_{i=1}^{N} \ln \left(\frac{1}{\Gamma(k)\vartheta^k} v_i^{k-1} \mathrm{e}^{-\frac{v_i}{\vartheta}} \right) && (2.4)\\ &= (k-1) \sum_{i=1}^{N} \ln v_i - \sum_{i=1}^{N} \frac{v_i}{\vartheta} - N\big(k \ln \vartheta + \ln \Gamma(k)\big). && (2.5) \end{aligned}$$

Now, in order to obtain the maximum likelihood estimate $\hat{\vartheta}$, we must differentiate the log-likelihood with respect to the estimand ϑ, and set it equal to

[1] This is possible because the values of ϑ which maximize the log-likelihood also maximize the likelihood.

zero:

$$\left.\frac{\partial \ell(v_1,\ldots,v_N;k,\vartheta)}{\partial \vartheta}\right|_{\vartheta=\hat{\vartheta}} = -\frac{kN}{\hat{\vartheta}} + \frac{\sum_{i=1}^N v_i}{\hat{\vartheta}^2} = 0 \tag{2.6}$$

$$\iff \hat{\vartheta} = \frac{\sum_{i=1}^N v_i}{kN} = \frac{\overline{v}}{k}\,. \tag{2.7}$$

This then allows us to find the maximum likelihood estimator $\widehat{\Theta}$, given by

$$\widehat{\Theta} = \frac{\sum_{i=1}^N V_i}{kN} = \frac{\overline{V}}{k}\,. \tag{2.8}$$

3. Asymptotic properties of the maximum likelihood estimator

We now wish to show some of the properties of this estimator. The definitions of these properties are given in Appendix A.

PROPERTY 3.1. *The maximum likelihood estimator derived in* (2.8) *is asymptotically unbiased, that is,*

$$\lim_{N\to+\infty} \mathbb{E}\left[g(V_1,\ldots,V_N);\vartheta\right] = \vartheta\,. \tag{3.1}$$

Proof. We wish to prove that $\lim_{N\to+\infty}\mathbb{E}\left[\frac{\overline{V}}{k}\right] = \vartheta$. We recall that $\mathbb{E}\left[V_i\right] = k\vartheta$ for $V_i \sim \Gamma(k,\vartheta)$ and that the expected value operator is linear to obtain

$$\mathbb{E}\left[\frac{\overline{V}}{k}\right] = \frac{\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N V_i\right]}{k} = \frac{\frac{1}{N}\sum_{i=1}^N \mathbb{E}\left[V_i\right]}{k} = \frac{\frac{1}{N}Nk\vartheta}{k} = \vartheta\,. \tag{3.2}$$

This proves that the maximum likelihood estimator of (2.8) is unbiased, hence it is also asymptotically unbiased. □

PROPERTY 3.2. *The maximum likelihood estimator derived in* (2.8) *is efficient.*

Proof. We use the fact that the random variables are independent to simplify the computations. Since ϑ is a scalar parameter, the Fisher information matrix is a scalar, equal to

$$\mathcal{I}(\vartheta) = -N\mathbb{E}\left[\frac{\partial^2}{\partial\vartheta^2}\left((k-1)\ln v_1 - \frac{v_1}{\vartheta} - \big(k\ln\vartheta + \ln\Gamma(k)\big)\right)\right] \tag{3.3}$$

$$= N\mathbb{E}\left[\frac{\partial^2}{\partial\vartheta^2}\Big(\frac{v_1}{\vartheta} + k\ln\vartheta\Big)\right] = \frac{kN}{\vartheta^2}\,. \tag{3.4}$$

We must also compute the variance of the ML estimator $\widehat{\Theta}$, which is given by

$$\mathbb{V}\left[\widehat{\Theta}\right] = \mathbb{V}\left[\frac{\overline{V}}{k}\right] = \frac{\vartheta^2}{kN}\,. \tag{3.5}$$

The Cramér–Rao lower bound is thus reached for all values of ϑ, which concludes the proof. $\square$

PROPERTY 3.3. *The maximum likelihood estimator of* (2.8) *is best asymptotically normal.*

Proof. In our case, we can show using the Cramér–Rao lower bound that Σ is minimal if it is equal to $\mathcal{I}^{-1}(\vartheta)$. To alleviate notations, we will write $\ell(\vartheta)$ instead of $\ell(v_1, \ldots, v_N; k, \vartheta)$. By definition, since $\hat{\vartheta} = \arg\max_{\vartheta} \ell(\vartheta)$, we know that $\ell'(\hat{\vartheta}) = 0$. Let ϑ_0 be the true value of the parameter ϑ. We can then use Taylor expansion on $\ell'(\hat{\vartheta})$ around $\hat{\vartheta} = \vartheta_0$ to obtain

$$\ell'(\hat{\vartheta}) = \ell'(\vartheta_0) + \frac{\ell''(\vartheta_0)}{1!}(\hat{\vartheta} - \vartheta_0) + \mathcal{O}\left((\hat{\vartheta} - \vartheta_0)^2\right) . \tag{3.6}$$

We know the expression on the left is zero, hence

$$\ell'(\vartheta_0) = -\ell''(\vartheta_0)(\hat{\vartheta} - \vartheta_0) + \mathcal{O}\left((\hat{\vartheta} - \vartheta_0)^2\right) . \tag{3.7}$$

Rearranging and multiplying by $\sqrt{n}$, we get

$$\sqrt{n}(\hat{\vartheta} - \vartheta_0) = \frac{\ell'(\vartheta_0)/\sqrt{n}}{-\ell''(\vartheta_0)/n + \mathcal{O}\left((\hat{\vartheta} - \vartheta_0)/n\right)} . \tag{3.8}$$

Next, we need to show that $\frac{1}{\sqrt{n}}\ell'(\vartheta_0) \sim \mathcal{N}\big(0, \mathcal{I}(\vartheta_0)\big)$. This is done using the Lindeberg–Lévy central limit theorem, in Appendix B. We know that $\frac{1}{N}\ell''(\vartheta_0) = \mathcal{I}(\vartheta_0)$. Finally, we can rewrite

$$\sqrt{N}(\hat{\vartheta} - \vartheta_0) \sim \frac{\mathcal{N}(0, \mathcal{I}(\vartheta_0))}{\mathcal{I}(\vartheta_0)} = \mathcal{N}(0, \mathcal{I}^{-1}(\vartheta_0)) , \tag{3.9}$$

where we didn't take into account the remainder of the Taylor series, which goes to zero. This proves that the ML estimator is best asymptotically normal. $\square$

PROPERTY 3.4. *The maximum likelihood estimator of* (2.8) *is consistent.*

Proof. We have shown that the estimator is unbiased, hence its MSE is equal to its variance. Since the estimator is efficient by Property 3.2, we know that its variance is equal to the Cramér–Rao lower bound, $\operatorname{cov}\widehat{\Theta} = \mathcal{I}^{-1}(\vartheta)$. We found this lower bound to be equal to $\frac{\vartheta^2}{kN}$ in (3.4). We have

$$\lim_{N \to +\infty} \operatorname{cov}\widehat{\Theta} = \lim_{N \to +\infty} \frac{\vartheta^2}{kN} = 0 . \tag{3.10}$$

This proves that the variance (and hence the mean square error) of the estimator goes to zero as N goes to infinity, hence the estimator is consistent. $\square$

4. Joint maximum likelihood estimation

We now consider $V_i \sim \Gamma(k,s)$ (for $i = 1,\dots,N$) with both k and s unknown. Before, we assumed k was known, so we could maximize the log-likelihood function with respect to s. Now, we have to maximize this function with respect to s and k at the same time. We know the maximum likelihood estimator of s, $\hat{s} = f(k)$. Therefore, in the log-likelihood function, we can replace all the occurrences of s by the estimator we found, $\hat{s}$. One then gets a function of k only, which can be differentiated and its derivative set to zero. Solving this, one can find the maximum likelihood estimator of k. We abusively write $\ell(v)$ instead of $\ell(v_1,\dots,v_N;\vartheta)$. Substituting in the estimator $\hat{s}$ instead of ϑ in the log-likelihood given in (2.5), one finds

$$\ell(v) = (k-1)\sum_{i=1}^{N}\ln v_i - \sum_{i=1}^{N}\frac{kv_i}{\overline{v}} - Nk\ln\overline{v} + Nk\ln k - N\ln\Gamma(k)\big)\,. \tag{4.1}$$

Taking the derivative of this function with respect to k, we get

$$\left.\frac{\partial\ell(v)}{\partial k}\right|_{k=\hat{k}} = \sum_{i=1}^{N}\ln v_i - N - N\ln\overline{v} + N\ln\hat{k} + \frac{N\hat{k}}{\hat{k}} - N\frac{\Gamma(\hat{k})}{\Gamma(\hat{k})}\psi_0(\hat{k}) \tag{4.2}$$

$$= \sum_{i=1}^{N}\ln v_i - N\ln\sum_{i=1}^{N}v_i + N\ln\hat{k} + N\ln N - N\psi_0(\hat{k})\,, \tag{4.3}$$

where ψ_0 is the digamma function, i.e. the logarithmic derivative of the gamma function. One must now look for a root of this equation:

$$\ln\hat{k} - \psi_0(\hat{k}) = \ln\left(\sum_{i=1}^{N}v_i\right) - \ln N - \frac{\sum_{i=1}^{N}\ln v_i}{N} \tag{4.4}$$

$$\iff \ln\hat{k} - \psi_0(\hat{k}) = \ln\left(\frac{\sum_{i=1}^{N}v_i}{N}\right) - \frac{\sum_{i=1}^{N}\ln v_i}{N}\,. \tag{4.5}$$

This equation has no closed-form solution for $\hat{k}$, but can be approximated using numerical methods since the function is very well-behaved.

5. Numerical experiments with various estimators

For the numerical simulation, N random variables were generated from a distribution with parameters $\Gamma(1,2)$, for different values of N (`10:50:1000`). For each value of N, the experiment was repeated $M = 500$ times.

In order to use method of moments estimation for the Gamma distribution given in (2.1), one first needs to know its characteristic function: $\varphi_{V_i}(t) = \mathbb{E}\left[e^{jtV_i}\right] = (1-jst)^{-k}$. The nth moment is given by $\mu_n = \mathbb{E}\left[V_i^n\right] = j^{-n}\varphi_{V_i}^{(n)}(0)$. Since the parameter vector has dimension two, we need to compute the first two moments. These are given by $\mu_1 = ks$ and $\mu_2 = k(k+1)s^2$. One can use the sample moments $\hat{\mu}_1 = \frac{1}{N}\sum_{i=1}^N v_i$ and $\hat{\mu}_2 = \frac{1}{N}\sum_{i=1}^N v_i^2$ to estimate μ_1 and μ_2. Using some simple algebra, one then finds

$$\hat{k}_{\mathrm{MOM}} = \frac{\hat{\mu}_1^2}{\hat{\mu}_2 - \hat{\mu}_1^2}\,, \quad \hat{s}_{\mathrm{MOM}} = \frac{\hat{\mu}_2}{\hat{\mu}_1} - \hat{\mu}_1\,. \tag{5.1}$$

The maximum likelihood estimators are computed using the formulas in (2.8) and (4.5), for the given sample. As mentioned in Section 4, the maximum likelihood estimator for k has no closed-form solution, but can be approximated using numerical methods which require an initial guess. One such first guess is provided by the method of moments estimator for the parameter, that is

$$\hat{k}_{\mathrm{ML}}^{(0)} = \hat{k}_{\mathrm{MOM}} = \frac{\hat{\mu}_1^2}{\hat{\mu}_2 - \hat{\mu}_1^2}\,. \tag{5.2}$$

Another possible choice for the first guess is

$$\hat{k}_{\mathrm{ML}}^{(0)} = \frac{3 - \xi + \sqrt{(\xi-3)^2 + 24\xi}}{12\xi}\,, \quad \text{where } \xi = \ln \overline{v} + \frac{1}{N}\sum_{i=1}^N \ln v_i\,. \tag{5.3}$$

This guess can be shown to be within 1.5 % of the actual value. The estimator for s can then be found from (2.8), using $\hat{k}_{\mathrm{ML}}$ instead of k.

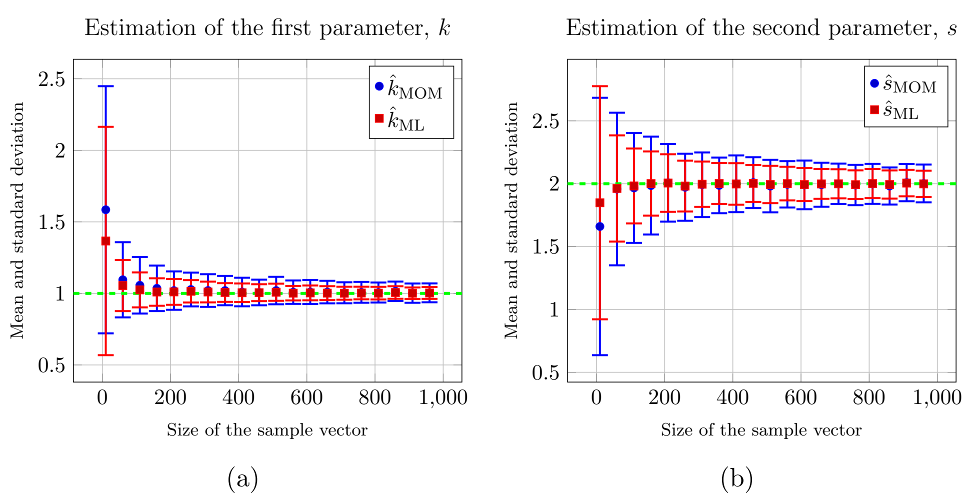

Figure 1. Both the method of moments estimator and the maximum likelihood estimator get increasingly accurate as the sample size goes up. The green line is the true value of the parameter, while the dots and squares indicate the mean of the estimators for a given value of N. The error bars are determined by the standard deviation of the estimators.

On Figures 1a and 1b, the mean and standard deviation are shown for the M values of both the method of moments estimator and the maximum likelihood estimator, for different values of the size of the sample, N. For both parameters, both estimators are unbiased, but the ML estimator has a lower variance. This should not come as a surprise: since it is efficient, every other estimator must have a greater or equal asymptotic variance. As is shown numerically in Section 7, this variance asymptotically goes to $\mathcal{I}^{-1}(\vartheta)$, the Cramér–Rao lower bound. This is also expected, in light of Property 3.2.[2]

6. **An analytical derivation of the Fisher information matrix**

One can analytically derive the Fisher information matrix. Since there are two estimators, the dimensions of the matrix are 2×2. Using the definition of the Fisher information matrix given in Theorem A.1, the Fisher matrix is given by

$$\mathcal{I}(\vartheta) = \begin{pmatrix} -\mathbb{E}\left[\dfrac{\partial^2 \ell(v)}{\partial s^2}\right] & -\mathbb{E}\left[\dfrac{\partial^2 \ell(v)}{\partial k \partial s}\right] \\ -\mathbb{E}\left[\dfrac{\partial^2 \ell(v)}{\partial s \partial k}\right] & -\mathbb{E}\left[\dfrac{\partial^2 \ell(v)}{\partial k^2}\right] \end{pmatrix}, \tag{6.1}$$

where $\ell(v)$ is used as a shorthand for $\ell(v_1, \ldots, v_N; k, s)$, given by

$$\ell(v_1, \ldots, v_N; k, s) = (k-1)\sum_{i=1}^{N} \ln v_i - \sum_{i=1}^{N} \frac{v_i}{s} - N\big(k \ln s + \ln \Gamma(k)\big). \tag{6.2}$$

Computing the entries of the matrix is fairly tedious, and the details are in Appendix C. The Fisher information matrix is then

$$\mathcal{I}(\vartheta) = N \begin{pmatrix} \frac{k}{s^2} & \frac{1}{s} \\ \frac{1}{s} & \psi_1(k) \end{pmatrix}, \tag{6.3}$$

and the Cramér–Rao lower bound, given by the inverse of the Fisher information matrix, is

$$\mathcal{I}^{-1}(\vartheta) = \frac{1}{N} \begin{pmatrix} \psi_1(k) & -\frac{1}{s} \\ -\frac{1}{s} & \frac{k}{s^2} \end{pmatrix} \frac{s^2}{k\psi_1(k) - 1}, \tag{6.4}$$

where ψ_1 is the trigamma function, i.e. the second derivative of the logarithm of the gamma function.

[2]While Property 3.2 is only proved for the estimator in (2.8), it also holds for the estimator found in Section 4.

7. Numerical evidence of convergence to the Cramér–Rao bound

Figure 2 gives the spectral norm of

$$\mathscr{R} = \operatorname{cov} \widehat{\Theta} \oslash \mathcal{I}(\vartheta) - \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \tag{7.1}$$

where $\oslash$ denotes Hadamard (element-wise) division and $\widehat{\Theta}$ is $(\hat{k}_{\mathrm{ML}}, \hat{s}_{\mathrm{ML}})$, for different values of N. If this estimator is efficient, we know by Theorem A.1 that its covariance should asymptotically go to the inverse of the Fisher information matrix. If one takes the Hadamard division of the covariance matrix by the inverse of the Fisher matrix, the result should be a matrix of ones. In order to visualize this convergence, the matrix must be "centered", by removing one in every position[3].

For any norm[4] of the resulting matrix $\mathscr{R}$, one can then observe that $\lim_{N\to+\infty} \|\mathscr{R}\| = 0$. This is shown for the spectral norm on Figure 2. To generate this picture, estimators were computed for samples of size $N \in \{ 10, 50, 150, 3000 \}$. This experiment was repeated $M = 10000$ times. The ratio matrix was then computed as the element-wise division between the empirical covariance of the estimators for a given N and the inverse of the Fisher information matrix, given by (6.4). Finally, the matrix was centered, and its spectral norm was computed.

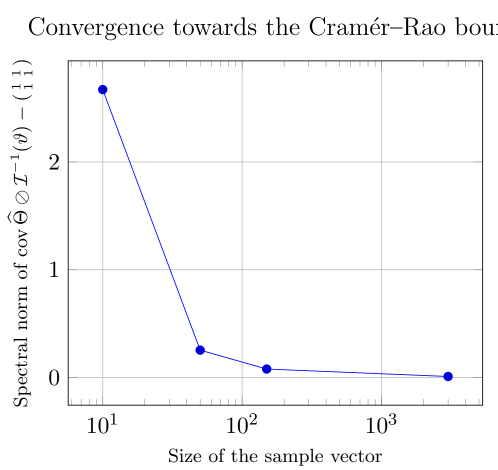

Figure 2. Using an adequately-centered spectral norm, one can visualize to which extent the Cramér–Rao lower bound is reached. The symbol "$\oslash$" denotes Hadamard division.

[3] If the matrix is not centered so as to make its asymptotic value equal to the zero matrix, then the norm does not suffice to prove convergence.

[4] For this paper, the spectral norm is used, but any norm would give similar results.

Appendix A. Definitions of properties

Definition A.1 (Unbiased estimator). The Fisher estimator $\widehat{\Theta} = g(Z)$ of ϑ is *unbiased* if

$$m_{\widehat{\Theta};\vartheta} \coloneqq \mathbb{E}\left[g(Z);\vartheta\right] = \vartheta\,, \quad \text{for all } \vartheta\,. \tag{A.1}$$

THEOREM A.1 (Cramér–Rao inequality). *If $Z = (Z_1, \ldots, Z_N)^T$ with i.i.d. random variables Z_k and if its probability density function given by $f_Z(z;\vartheta) = \prod_{k=1}^{N} f_{Z_k}(z_k;\vartheta)$ satisfies certain regularity conditions, then the covariance of any unbiased estimator $\widehat{\Theta}$ satisfies the* Cramér–Rao inequality

$$\operatorname{cov}\widehat{\Theta} \succeq \mathcal{I}^{-1}(\vartheta)\,, \tag{A.2}$$

where $\mathcal{I}(\vartheta)$ is the $p \times p$ Fisher information matrix, *defined by*

$$\big[\mathcal{I}(\vartheta)\big]_{i,j} \coloneqq -\mathbb{E}\left[\frac{\partial^2 \ln f_Z(z;\vartheta)}{\partial\vartheta_i \partial\vartheta_j}\right]\,. \tag{A.3}$$

Definition A.2 (Efficient estimator). An unbiased estimator is said to be *efficient* if it reaches the Cramér–Rao bound for all values of ϑ, that is,

$$\operatorname{cov}\widehat{\Theta} = \mathcal{I}^{-1}(\vartheta)\,, \quad \forall\vartheta\,. \tag{A.4}$$

Definition A.3 (Best asymptotically normal estimator). A sequence $\{\widehat{\Theta}_N(Z)\}_{N\in\mathbb{N}}$ of consistent estimators of ϑ is called *best asymptotically normal* if

$$\sqrt{N}\left(\widehat{\Theta}_N(Z) - \vartheta\right) \xrightarrow[N\to+\infty]{\mathcal{D}} \mathcal{N}(0,\Sigma)\,, \tag{A.5}$$

for some minimal positive definite matrix Σ.

Definition A.4 (Consistent estimator). A sequence $\{\widehat{\Theta}_N(Z)\}_{N\in\mathbb{N}}$ of estimators of ϑ is called *consistent* if

$$\operatorname*{plim}_{N\to+\infty} \widehat{\Theta}_N(Z) = \vartheta\,. \tag{A.6}$$

Equivalently, one can show that the MSE of the estimator converges to zero as N goes to infinity.

Appendix B. Omitted proofs

The following is a proof of normality, using the Lindeberg–Lévy formulation of the central limit theorem.

Proof. We want to prove that $\frac{1}{\sqrt{N}}\ell'(\vartheta_0) \sim \mathcal{N}\big(0, \mathcal{I}(\vartheta_0)\big)$. We simplify notation by writing $f(v_i)$ instead of $f_{V_i}(v_i;k,\vartheta)$ and using Euler's notation for derivatives. First, we show that the expected value of $\frac{1}{\sqrt{N}}\ell'(\vartheta_0)$ is zero.

$$\mathbb{E}\left[\frac{\ell'(\vartheta_0)}{\sqrt{N}}\right] = \int_{-\infty}^{+\infty} \partial_{\vartheta_0}\left(\sum_{i=1}^{N}\frac{\ln f(v_i)}{\sqrt{N}}\right) f(v_i)\,\mathrm{d}v_i \tag{B.1}$$

$$= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\int_{-\infty}^{+\infty}\frac{\partial_\vartheta f(v_i)}{f(v_i)} f(v_i)\,\mathrm{d}v_i \tag{B.2}$$

$$= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\int_{-\infty}^{+\infty}\frac{\partial_\vartheta f(v_i)}{f(v_i)} f(v_i)\,\mathrm{d}v_i \tag{B.3}$$

$$= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\int_{-\infty}^{+\infty}\partial_\vartheta f(v_i)\,\mathrm{d}v_i \tag{B.4}$$

$$= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\partial_\vartheta\int_{-\infty}^{+\infty} f(v_i)\,\mathrm{d}v_i \tag{B.5}$$

$$= 0\,. \tag{B.6}$$

Next, we compute the variance of $\frac{1}{N}\ell'(\vartheta_0)$ (for $i = 1, \dots, N$)

$$\mathbb{E}\left[\left(\partial_{\vartheta_0}\ln f(v_i)\right)^2\right] = \int_{-\infty}^{+\infty}\partial_{\vartheta_0}\ln f(v_i)\frac{\partial_{\vartheta_0} f(v_i)}{f(v_i)} f(v_i)\,\mathrm{d}v_i \tag{B.7}$$

$$= \int_{-\infty}^{+\infty}\partial_{\vartheta_0}\ln f(v_i)\partial_{\vartheta_0} f(v_i)\,\mathrm{d}v_i\,. \tag{B.8}$$

Using the product rule, we can then find

$$\begin{aligned} &= -\int_{-\infty}^{+\infty}\partial_{\vartheta_0\vartheta_0}\ln f(v_i) f(v_i)\,\mathrm{d}v_i \\ &\quad + \int_{-\infty}^{+\infty}\partial_{\vartheta_0}\left(\partial_{\vartheta_0}\ln f(v_i) f(v_i)\right)\mathrm{d}v_i \end{aligned} \tag{B.9}$$

$$= -\mathbb{E}\left[\partial_{\vartheta_0\vartheta_0}\ln f(v_i)\right] + \partial_{\vartheta_0}\int_{-\infty}^{\infty}\frac{\partial_{\vartheta_0} f(v_i)}{f(v_i)} f(v_i)\,\mathrm{d}v_i \tag{B.10}$$

$$= \mathcal{I}(\vartheta)\,, \tag{B.11}$$

where the last expression can be shown to be zero by a similar argument as the one used above for the expected value. Knowing this, one easily finds

$$\mathbb{V}\left[\frac{\ell'(\vartheta_0)}{\sqrt{N}}\right] = \frac{1}{N}\mathbb{V}\left[\sum_{i=1}^{N}\partial_{\vartheta_0}\ln f(v_i)\right] = \mathcal{I}(\vartheta_0)\,, \tag{B.12}$$

since the random variables are i.i.d.. Using the Lindeberg–Lévy CLT, we thus have

$$\frac{\ell'(\vartheta_0)}{\sqrt{N}} \sim \mathcal{N}(0, \mathcal{I}(\vartheta_0))\,. \tag{B.13}$$

□

Appendix C. Computation of the Fisher information matrix

In order to derive an analytical expression for the Fisher information matrix, we start by computing various derivatives of the log-likelihood which will be needed later (writing $\ell(v)$ instead of $\ell(v_1, \ldots, v_N; k, s)$):

$$\frac{\partial \ell(v)}{\partial s} = \frac{1}{s^2}\sum_{i=1}^{N} v_i - \frac{Nk}{s}\,, \tag{C.1}$$

$$\frac{\partial \ell(v)}{\partial k} = \sum_{i=1}^{N} \ln v_i - N\ln s - N\psi_0(k)\,, \tag{C.2}$$

$$\frac{\partial^2 \ell(v)}{\partial s^2} = -\frac{2}{s^3}\sum_{i=1}^{N} v_i + \frac{Nk}{s^2}\,, \tag{C.3}$$

$$\frac{\partial^2 \ell(v)}{\partial k^2} = -N\psi_1(k)\,, \tag{C.4}$$

$$\frac{\partial^2 \ell(v)}{\partial k \partial s} = \frac{\partial^2 \ell(v)}{\partial s \partial k} = -\frac{N}{s}\,. \tag{C.5}$$

In order to find the entries of the Fisher matrix, one must take the expectation of the previous derivatives:

$$\Big[\mathcal{I}(\vartheta)\Big]_{0,0} = -\mathbb{E}\left[\frac{\partial^2 \ell(v)}{\partial s^2}\right] = -\mathbb{E}\left[-\frac{2}{s^3}\sum_{i=1}^{N} v_i + \frac{Nk}{s^2}\right] \tag{C.6}$$

$$= \frac{2}{s^3}Nks - \frac{Nk}{s^2} = \frac{Nk}{s^2}\,, \tag{C.7}$$

$$\Big[\mathcal{I}(\vartheta)\Big]_{0,1} = \Big[\mathcal{I}(\vartheta)\Big]_{1,0} = -\mathbb{E}\left[\frac{\partial^2 \ell(v)}{\partial k \partial s}\right] = \frac{N}{s}\,, \tag{C.8}$$

$$\Big[\mathcal{I}(\vartheta)\Big]_{1,1} = -\mathbb{E}\left[\frac{\partial^2 \ell(v)}{\partial k^2}\right] = -N\psi_1(k)\,. \tag{C.9}$$

References

Université catholique de Louvain, Ottignies-Louvain-la-Neuve, Belgium
E-mail: navarre.louis@student.uclouvain.be

Université catholique de Louvain, Ottignies-Louvain-la-Neuve, Belgium
E-mail: gilles.peiffer@student.uclouvain.be