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. Maximum likelihood estimation

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 $\theta \coloneqq s$ 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 $\mathcal{L}(v_1, \ldots, v_N; k, \theta)$ is given by

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

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

$$\ell(v_1, \ldots, v_N; k, \theta) \coloneqq \ln \mathcal{L}(v_1, \ldots, v_N; k, \theta) \tag{2.3}$$

$$= \sum_{i=1}^{N} \ln \left(\frac{1}{\Gamma(k)\theta^k} v_i^{k-1} \mathrm{e}^{-\frac{v_i}{\theta}} \right) \tag{2.4}$$

$$= (k-1) \sum_{i=1}^{N} \ln v_i - \sum_{i=1}^{N} \frac{v_i}{\theta} - N\Big(k \ln \theta + \ln \Gamma(k)\Big) . \tag{2.5}$$

Now, in order to obtain the maximum likelihood estimate $\hat{\theta}$, 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:

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

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

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. Properties of the estimator

The estimator we just found has many interesting properties. First, it is an asymptotically unbiased estimator (A.1) . One can show that $lim_{N\to+\infty}\mathbb{E}\left[\frac{\overline{V}}{k}\right] = \theta$:

$$\mathbb{E}\left[\frac{\overline{V}}{k}\right] = \frac{\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N V_i\right]}{k} = \frac{\frac{1}{N}\sum_{i=1}^N \mathbb{E}\left[V_i\right]}{k} = \frac{\frac{1}{N}Nk\theta}{k} = \theta\,. \tag{3.1}$$

The estimator is more than asymptotically unbiased; it is unbiased. Another interesting property is the efficiency (A.2) of θ. An unbiased estimator is said to be *efficient* if it reaches the Cramr-Rao bound for all values of θ, that is,

$$\operatorname{cov}\widehat{\Theta} = \mathcal{I}^{-1}(\theta)\,, \quad \forall\theta\,. \tag{3.2}$$

Since we only have one estimator, we only have to calculate the expectation of the second derivative of the log-likelihood function derived earlier:

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

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

Then, we have to compute the variance of the ML estimator $\hat\Theta$, which is given by

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

We directly see that the variance of our estimator reaches the Cramr-Rao lower bound; it is then an efficient estimator.

A third property is that the estimator is best asymptotically normal (A.3). In our case, we can show using the Cramr–Rao lower bound that Σ is minimal if it is equal to $\mathcal{I}^{-1}(\theta)$. To alleviate notations, we will write $\ell(\theta)$ instead of $\ell(v_1,\ldots,v_N;k,\theta)$. By definition, since $\hat\theta = \arg\max_\theta \ell(\theta)$, we know that

$\ell'(\hat{\theta}) = 0$. Let θ_0 be the true value of the parameter θ. We can then use Taylor expansion on $\ell'(\hat{\theta})$ around $\hat{\theta} = \theta_0$ to obtain

$$\ell'(\hat{\theta}) = \ell'(\theta_0) + \frac{\ell''(\theta_0)}{1!}(\hat{\theta} - \theta_0) + \mathcal{O}\left((\hat{\theta} - \theta_0)^2\right) . \tag{3.6}$$

We know the expression on the left is zero, hence

$$\ell'(\theta_0) = -\ell''(\theta_0)(\hat{\theta} - \theta_0) + \mathcal{O}\left((\hat{\theta} - \theta_0)^2\right) . \tag{3.7}$$

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

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{\ell'(\theta_0)/\sqrt{n}}{-\ell''(\theta_0)/n + \mathcal{O}\left((\hat{\theta} - \theta_0)/n\right)} . \tag{3.8}$$

Next, we need to show that $\ell'(\theta_0)/\sqrt{n} \sim \mathcal{N}\big(0, \mathcal{I}(\theta_0)\big)$. This is done using the Lindeberg–Lvy central limit theorem, in Appendix B.1. We know that $1/N\ell''(\theta_0) = \mathcal{I}(\theta_0)$. Finally, we can rewrite

$$\sqrt{N}(\hat{\theta} - \theta_0) \sim \frac{\mathcal{N}(0, \mathcal{I}(\theta_0))}{\mathcal{I}(\theta_0)} = \mathcal{N}(0, \mathcal{I}^{-1}(\theta_0)) , \tag{3.9}$$

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.

A last but important property to show that we have a "good" estimator is the consistency (A.4) of it. We have shown that the estimator is unbiased, hence its MSE is equal to its variance. Since the estimator is efficient - as said earlier - we know that its variance is equal to the Cramr–Rao lower bound, $\operatorname{cov}\widehat{\Theta} = \mathcal{I}^{-1}(\theta)$. We found this lower bound to be equal to $\theta^2/(kN)$ in (3.4). We have

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

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.

4. Joint maximum likelihood estimation

We now consider $V_i \sim \Gamma(k, s)$ (for $i = 1, \ldots, N$) 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, $\hat{s}$. We then get a function of k only, which we can differentiate and set its derivative to zero, solving to find the maximum likelihood estimator of k. We use the log-likelihood as given in (2.5). We abusively write $\ell(v)$ instead of $\ell(v_1, \ldots, v_N; \theta)$ Substituting in the estimator $\hat{s}$ instead of s, one finds

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

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

$$\left.\frac{\partial \ell(v)}{\partial k}\right|_{k=\hat{k}} = \sum_{i=1}^{N} \ln v_i - N - N\ln\overline{v} + N\ln\hat{k} + \frac{N\hat{k}}{\hat{k}} - N\frac{\Gamma(\hat{k})}{\Gamma(\hat{k})}\psi^{(0)}(\hat{k}) \tag{4.2}$$

$$= \sum_{i=1}^{N} \ln v_i - N\ln\sum_{i=1}^{N} v_i + N\ln\hat{k} + N\ln N - N\psi^{(0)}(\hat{k}) , \tag{4.3}$$

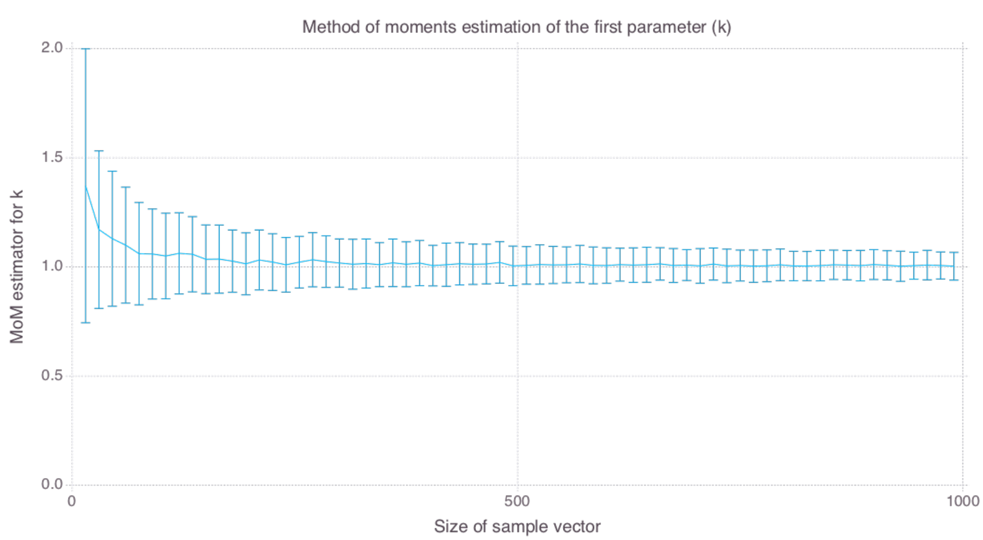
where $\psi^{(0)}$ is the digamma function, i.e. the logarithmic derivative of the gamma function. We now look for a root of this equation:

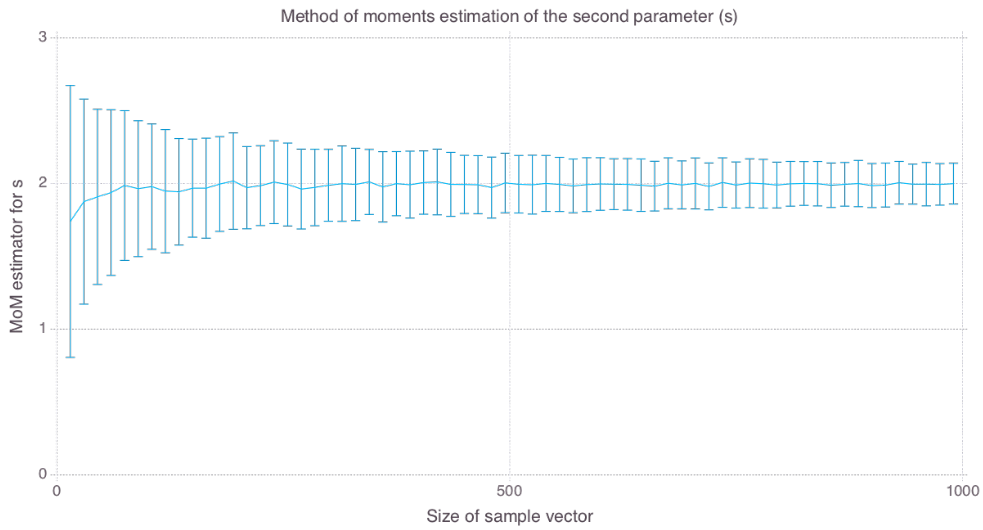
$$\ln\hat{k} - \psi^{(0)}(\hat{k}) = \ln\left(\sum_{i=1}^{N} v_i\right) - \ln N - \frac{\sum_{i=1}^{N}\ln v_i}{N} \tag{4.4}$$

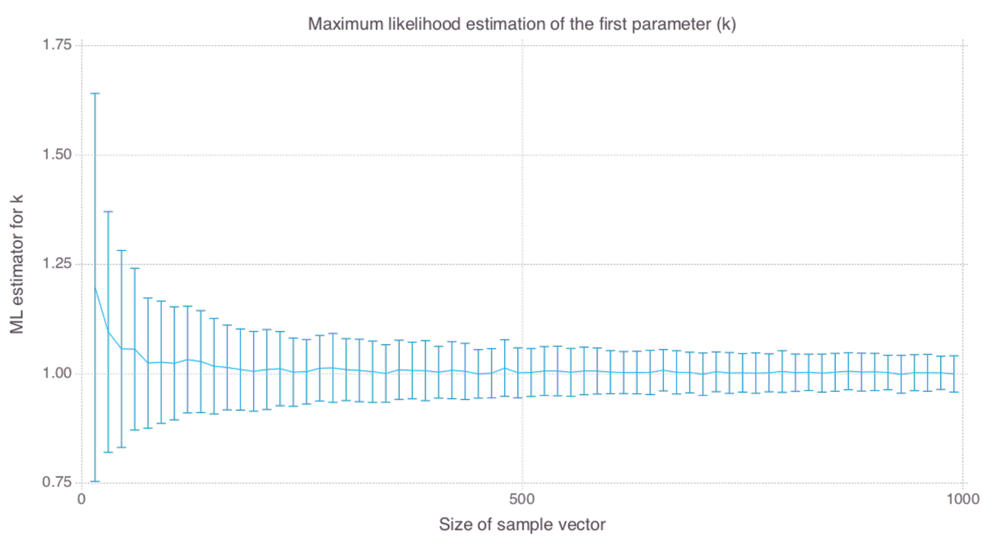
$$\iff \ln\hat{k} - \psi^{(0)}(\hat{k}) = \ln\left(\frac{\sum_{i=1}^{N} v_i}{N}\right) - \frac{\sum_{i=1}^{N}\ln v_i}{N} . \tag{4.5}$$

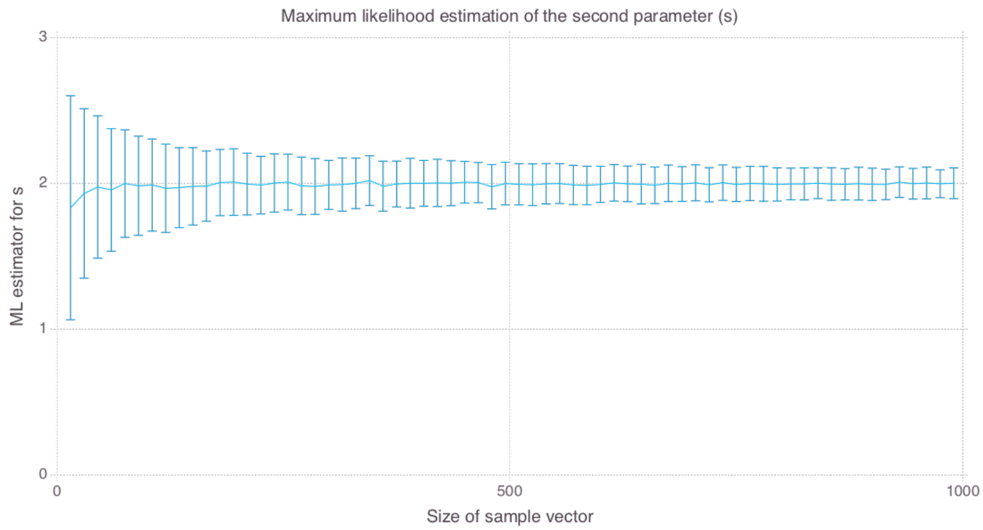
This equation has no closed-form solution for $\hat{k}$, but can be approximated using numerical methods since the function is very well-behaved.

5. Numerical simulation




Method of moments estimation of the second parameter (s)
MoM estimator for s
0
1
2
3
0
500
1000
Size of sample vector


Maximum likelihood estimation of the first parameter (k)
ML estimator for k
0.75
1.00
1.25
1.50
1.75
0
500
1000
Size of sample vector



6. Fisher information matrix

We can compute the Fisher information matrix. Since we have two estimators, this matrix is a 2×2 matrix. This matrix is hence computed by

$$\mathcal{I}(\theta) = \begin{pmatrix} -\mathbb{E}\left[\frac{\partial^2 l(x)}{\partial s^2}\right] & -\mathbb{E}\left[\frac{\partial^2 l(x)}{\partial k \partial s}\right] \\ -\mathbb{E}\left[\frac{\partial^2 l(x)}{\partial s \partial k}\right] & -\mathbb{E}\left[\frac{\partial^2 l(x)}{\partial k^2}\right] \end{pmatrix}$$

And remember the log-likelihood function, computed before.

$$\ell(x) = (k-1)\sum_{i=1}^{N} \ln v_i - \sum_{i=1}^{N} \frac{v_i}{s} - N\Big(k \ln s + \ln \Gamma(k)\Big) \tag{6.1}$$

The calculation of theses entries can be found in appendix C. Finally, the Fisher information matrix is

$$\mathcal{I}(\theta) = N \begin{pmatrix} \frac{K}{s^2} & \frac{1}{s} \\ \frac{1}{s} & \psi_1(k) \end{pmatrix}$$

And the inverse of this matrix, the Cramr-Rao lower-bound, is

$$\mathcal{I}^{-1}(\theta) = \frac{1}{N} \begin{pmatrix} \psi_1(k) & -\frac{1}{s} \\ -\frac{1}{s} & \frac{K}{s^2} \end{pmatrix} \cdot \frac{s^2}{K\psi_1(k) - 1}$$

7. Numerical proof

Figure 7 show the matrix norm induced by the vector 2-norm of the *ratio* matrix. This matrix is is the result of the ratio of the entries of the empirical covariance matrix by the inverse of the Fisher information matrix, subtracted by one. As a result, if the empirical covariance matrix tends to reach the

Cramr-Rao lower bound, the ratio of the entries should tend to 1, and hence the matrix induced norm of this matrix subtracted by 1 should be equal to 0. Indeed, we see on figure 7 that this matrix induced norm tends to 0 as the size of the sample vector increases. It empirically proofs that our estimators reach the Cramr-Rao lower bound, meaning that they are efficient.

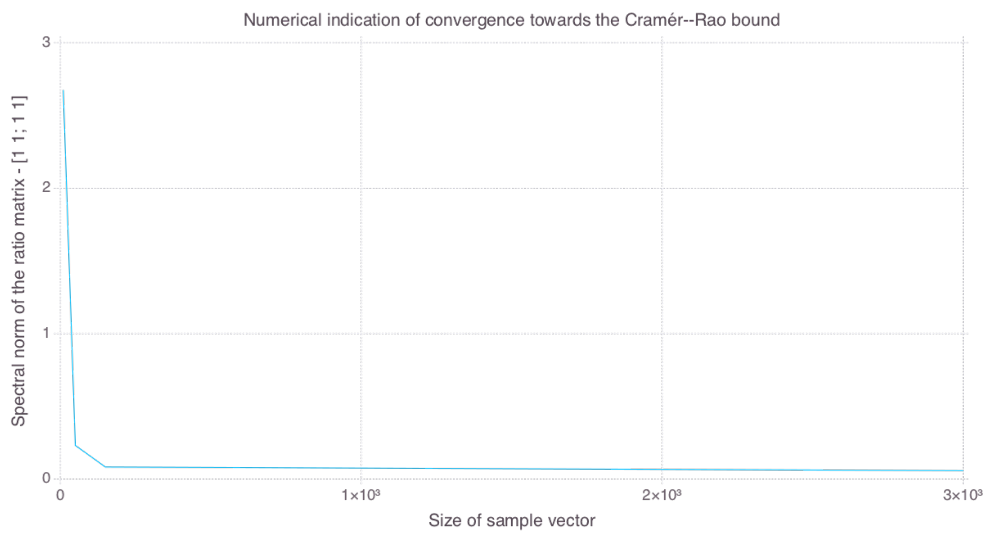


Figure 1. Norm of the "ratio" matrix subtracted by 1

Appendix A. Definitions of properties

A.1. *Asymptotically unbiased.*

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

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

A.2. *Efficiency.*

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

$$\operatorname{cov}\widehat{\Theta} = \mathcal{I}^{-1}(\theta)\,, \quad \forall\theta\,. \tag{A.2}$$

A.3. *Best asymptotically normal.*

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

$$\sqrt{N}\left(\widehat{\Theta}_N(Z)-\theta\right)\xrightarrow[N\to+\infty]{\mathcal{D}}\mathcal{N}(0,\Sigma)\,, \tag{A.3}$$

for some minimal positive definite matrix Σ.

A.4. *Consistency.*

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

$$\operatorname*{plim}_{N\to+\infty}\widehat{\Theta}_N(Z)=\theta\,. \tag{A.4}$$

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

Appendix B. Omitted proofs

B.1. *Proof of normality using the central limit theorem.*

Proof. We want to prove that $\ell'(\theta_0)/\sqrt{N}\sim\mathcal{N}\big(0,\mathcal{I}(\theta_0)\big)$. We simplify notation by writing $f(v_i)$ instead of $f_{V_i}(v_i;k,\theta)$ and using Euler's notation for derivatives. First, we show that the expected value of $\ell'(\theta_0)/\sqrt{N}$ is zero.

$$\mathbb{E}\left[\frac{\ell'(\theta_0)}{\sqrt{N}}\right]=\int_{-\infty}^{+\infty}\partial_{\theta_0}\left(\sum_{i=1}^{N}\frac{\ln f(v_i)}{\sqrt{N}}\right)f(v_i)\,\mathrm{d}v_i \tag{B.1}$$

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

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

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

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

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

Next, we compute the variance of $\ell'(\theta_0)/\sqrt{N}$. First, we compute (for $i = 1, \ldots, N$)

$$\mathbb{E}\left[\left(\partial_{\theta_0} \ln f(v_i)\right)^2\right] = \int_{-\infty}^{+\infty} \partial_{\theta_0} \ln f(v_i) \frac{\partial_{\theta_0} f(v_i)}{f(v_i)} f(v_i)\, \mathrm{d}v_i \tag{B.7}$$

$$= \int_{-\infty}^{+\infty} \partial_{\theta_0} \ln f(v_i) \partial_{\theta_0} f(v_i)\, \mathrm{d}v_i\,. \tag{B.8}$$

Using the product rule, we can then find

$$\begin{aligned} = \; & - \int_{-\infty}^{+\infty} \partial_{\theta_0\theta_0} \ln f(v_i) f(v_i)\, \mathrm{d}v_i \\ & + \int_{-\infty}^{+\infty} \partial_{\theta_0}\left(\partial_{\theta_0} \ln f(v_i) f(v_i)\right) \mathrm{d}v_i\,. \end{aligned} \tag{B.9}$$

$$= -\mathbb{E}\left[\partial_{\theta_0\theta_0} \ln f(v_i)\right] + \partial_{\theta_0} \int_{-\infty}^{\infty} \frac{\partial_{\theta_0} f(v_i)}{f(v_i)} f(v_i)\, \mathrm{d}v_i \tag{B.10}$$

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

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

$$\mathbb{V}\left[\frac{\ell'(\theta_0)}{\sqrt{N}}\right] = \frac{1}{N} \mathbb{V}\left[\sum_{i=1}^{N} \partial_{\theta_0} \ln f(v_i)\right] = \mathcal{I}(\theta_0)\,, \tag{B.12}$$

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

$$\frac{\ell'(\theta_0)}{\sqrt{N}} \sim \mathcal{N}(0, \mathcal{I}(\theta_0))\,. \tag{B.13}$$

□

Appendix C. Calculations of the Fisher information matrix

$$\frac{\partial h(z;\theta)}{\partial s} = \frac{1}{s^2} \sum_{i=1}^{N} v_i - \frac{Nk}{s} \tag{C.1}$$

$$\frac{\partial h(z;\theta)}{\partial k} = \sum_{i=1}^{N} \ln v_i - N \ln s - N\psi_0(k) \tag{C.2}$$

$$\frac{\partial^2 h(z;\theta)}{\partial s^2} = -\frac{2}{s^3} \sum_{i=1}^{N} v_i + \frac{Nk}{s^2} \tag{C.3}$$

$$\frac{\partial^2 h(z;\theta)}{\partial k^2} = -N\psi_1(k). \tag{C.4}$$

$$\frac{\partial^2 h(z;\theta)}{\partial k \partial s} = -\frac{N}{s}. \tag{C.5}$$

Then, we take the expectation of these computed values.

$$\mathcal{I}_{00} = -\mathbb{E}\left\{\frac{\partial^2 h(z;\theta)}{\partial s^2}\right\} = -\mathbb{E}\left\{-\frac{2}{s^3}\sum_{i=1}^{N} v_i + \frac{Nk}{s^2}\right\} \tag{C.6}$$

$$= \frac{2}{s^3}Nsk - \frac{NK}{s^2} = \frac{NK}{s^2}. \tag{C.7}$$

$$\mathcal{I}_{01} = \mathcal{I}_{10} = -\mathbb{E}\left\{\frac{\partial^2 h(z;\theta)}{\partial k \partial s}\right\} = \frac{N}{s}. \tag{C.8}$$

$$\mathcal{I}_{11} = -\mathbb{E}\left\{\frac{\partial^2 h(z;\theta)}{\partial k^2}\right\} = -N\psi_1(k). \tag{C.9}$$

References

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

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