

# LINMA1731 – Project 2019

## Fish schools tracking

By LOUIS NAVARRE and GILLES PEIFFER

### Abstract

In this paper we propose a solution to the project for the class “Stochastic processes: Estimation and prediction” given during the Fall term of 2019 at the EPL by Luc Vandendorpe and Pierre-Antoine Absil. 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; numerical simulations are also included, using Julia and J<sup>u</sup>MP [1]. This paper also details how to obtain these speed measurements from noisy observations of the position. Using a particle filter, one can then try to use these noisy observations in order to obtain a more accurate position for each of the fish. Finally, the effect of various parameters of the model, such as the number of particles per fish, the sampling period and the standard deviation of the observational noise is studied.

### Contents

<b>Part 1. Average speed estimation</b>	2
1. Introduction	2
2. A maximum likelihood estimator for the scale parameter	2
3. Asymptotic properties of the maximum likelihood estimator	2
4. Joint maximum likelihood estimation	4
5. Numerical experiments with various estimators	4
6. An analytical derivation of the Fisher information matrix	6
7. Numerical evidence of convergence to the Cramér–Rao bound	6
<b>Part 2. Particle filtering</b>	7
8. Introduction	7
9. Estimation of the parameters from noisy measurements	7
10. Sequential Monte Carlo method	8
11. Mean squared error analysis	9
12. Conclusion	10
Appendix A. Definitions of properties	11
Appendix B. Omitted proofs	11
Appendix C. Computation of the Fisher information matrix	12
References	13

## 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 [4]. 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 the scale parameter

Let  $v_i$  be i.i.d. realisations of a random variable following a Gamma distribution  $\Gamma(k, s)$  (with  $i = 1, \dots, N$ ) [10]. 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 [13]. Since the estimand  $\vartheta$  is a deterministic quantity, we use Fisher estimation. In order to do this, let us restate the probability density function of  $V_i \sim \Gamma(k, s)$ :

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

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

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

In order to alleviate notation, we compute instead the log-likelihood, which is generally easier to work with<sup>1</sup>:

$$(2.3) \quad \ell(v_1, \dots, v_N; k, \vartheta) := \ln \mathcal{L}(v_1, \dots, v_N; k, \vartheta)$$

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

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

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

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

$$(2.7) \quad \iff \hat{\vartheta} = \frac{\sum_{i=1}^N v_i}{kN} = \frac{\bar{v}}{k}.$$

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

$$(2.8) \quad \hat{\Theta} = \frac{\sum_{i=1}^N V_i}{kN} = \frac{\bar{V}}{k}.$$

### 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.

---

<sup>1</sup>This is possible because the values of  $\vartheta$  which maximize the log-likelihood also maximize the likelihood.

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

$$(3.1) \quad \lim_{N \rightarrow +\infty} \mathbb{E}[g(V_1, \dots, V_N); \vartheta] = \vartheta.$$

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

$$(3.2) \quad \mathbb{E}\left[\frac{\bar{V}}{k}\right] = \frac{\mathbb{E}\left[\frac{1}{N} \sum_{i=1}^N V_i\right]}{k} = \frac{\frac{1}{N} \sum_{i=1}^N \mathbb{E}[V_i]}{k} = \frac{\frac{1}{N} N k \vartheta}{k} = \vartheta.$$

This proves that the maximum likelihood estimator of (2.8) is unbiased, hence it is also asymptotically unbiased. This can be seen on Figures 1a and 1b.  $\square$

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  $\vartheta$  is a scalar parameter, the Fisher information matrix is a scalar, equal to

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

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

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

$$(3.5) \quad \mathbb{V}[\widehat{\Theta}] = \mathbb{V}\left[\frac{\bar{V}}{k}\right] = \frac{\vartheta^2}{kN}.$$

The Cramr–Rao lower bound is thus reached for all values of  $\vartheta$ , which concludes the proof. This is shown on Figure 2.  $\square$

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

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

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

We know the expression on the left is zero, hence

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

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

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

Next, we need to show that  $\frac{1}{\sqrt{n}}\ell'(\vartheta_0) \sim \mathcal{N}(0, \mathcal{I}(\vartheta_0))$ . This is done using the Lindeberg–Lvy central limit theorem, in Appendix B [8]. We know that  $\frac{1}{N}\ell''(\vartheta_0) = \mathcal{I}(\vartheta_0)$ . Finally, we can rewrite

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

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 Cramr–Rao lower bound,  $\text{cov } \widehat{\Theta} = \mathcal{I}^{-1}(\vartheta)$ . We found this lower bound to be equal to  $\frac{\vartheta^2}{kN}$  in (3.4). We have

$$(3.10) \quad \lim_{N \rightarrow +\infty} \text{cov } \widehat{\Theta} = \lim_{N \rightarrow +\infty} \frac{\vartheta^2}{kN} = 0.$$

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  $\ell(\vartheta)$  instead of  $\ell(v_1, \dots, v_N; \vartheta)$ . Substituting in the estimator  $\hat{s}$  instead of  $\vartheta$  in the log-likelihood given in (2.5), one finds

$$(4.1) \quad \ell(\vartheta) = (k-1) \sum_{i=1}^N \ln v_i - \sum_{i=1}^N \frac{kv_i}{\bar{v}} - Nk \ln \bar{v} + Nk \ln k - N \ln \Gamma(k).$$

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

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

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

where  $\psi_0$  is the digamma function, i.e. the logarithmic derivative of the gamma function [2]. One must now look for a root of this equation:

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

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

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:  $\varphi_{V_i}(t) = \mathbb{E} [e^{itV_i}] = (1 - jst)^{-k}$ . The  $n$ th moment is given by  $\mu_n = \mathbb{E} [V_i^n] = j^{-n} \varphi_{V_i}^{(n)}(0)$  [9]. 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  $\mu_1$  and  $\mu_2$ . Using some simple algebra, one then finds

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

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 could be provided by the method of moments estimator for the parameter, that is

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

Another possible choice for the first guess is

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

This guess can be shown to be within 1.5 % of the actual value [5]. The estimator for  $s$  can then be found from (2.8), using  $\hat{k}_{\text{ML}}$  instead of  $k$ .

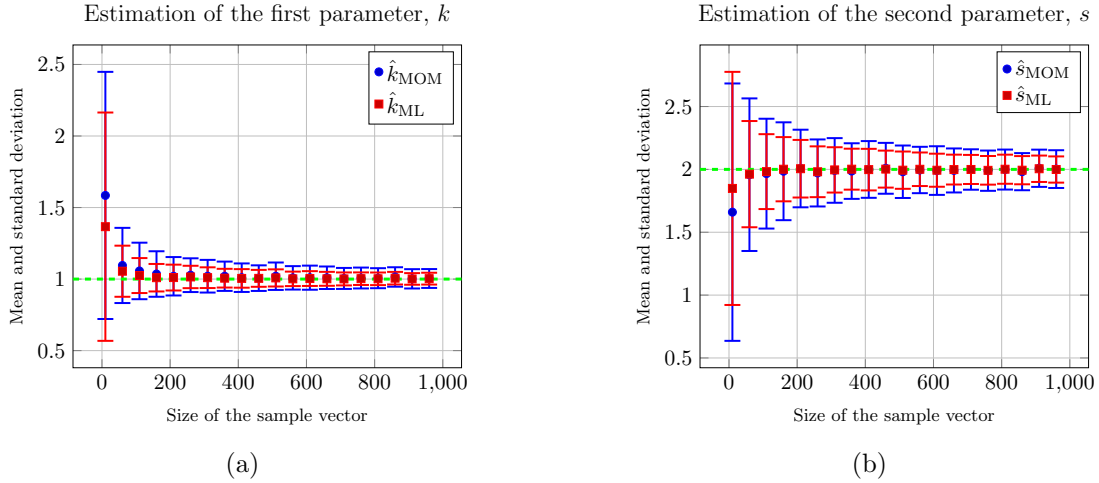


Figure 1. Both the method of moments estimator and the maximum likelihood estimator get increasingly accurate as the sample size goes up. The green line is the true value of the parameter, while the dots and squares indicate the mean of the estimators for a given value of  $N$ . The error bars are determined 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 Cramr–Rao lower bound. This is also expected, in light of Property 3.2.<sup>2</sup>

<sup>2</sup>While Property 3.2 is only proved for the estimator in (2.8), it also holds for the estimator found in (4.5), as it is a general property of maximum likelihood estimators.

## 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  $2 \times 2$ . Using the definition of the Fisher information matrix given in Theorem A.1, the Fisher matrix is given by

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

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

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

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

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

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

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

where  $\psi_1$  is the trigamma function, i.e. the second derivative of the logarithm of the gamma function [3].

## 7. Numerical evidence of convergence to the Cramr–Rao bound

Figure 2 gives the spectral norm as defined in [12] of

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

where  $\oslash$  denotes Hadamard division as defined in [11] and  $\widehat{\Theta}$  is  $(\hat{k}_{\text{ML}}, \hat{s}_{\text{ML}})$ , for different values of  $N$ . If this estimator is efficient, we know by Definition A.2 that its covariance should asymptotically go to the inverse of the Fisher information matrix.

Let the covariance matrix be defined as in [7], that is

$$(7.2) \quad \text{cov} \widehat{\Theta} = \frac{1}{M-1} \begin{pmatrix} \sum_{i=1}^M \hat{s}_{i,c}^2 & \sum_{i=1}^M \hat{k}_{i,c} \hat{s}_{i,c} \\ \sum_{i=1}^M \hat{s}_{i,c} \hat{k}_{i,c} & \sum_{i=1}^M \hat{k}_{i,c}^2 \end{pmatrix}.$$

where  $\hat{k}_i$  and  $\hat{s}_i$  are the maximum likelihood estimators for the sample in the  $i$ th repetition of the experiment, and  $\hat{\varrho}_{i,c} = \hat{\varrho}_i - \bar{\varrho}$ , with  $\bar{\varrho}$  denoting the mean over all experiments of  $\hat{\varrho}$  for a given value of  $N$ , for  $\varrho = k, s$ .

If one takes the Hadamard division of the covariance matrix by the inverse of the Fisher information matrix, the result should converge to a matrix of ones [11]. In order to visualize this convergence, the matrix must be “centered”, by removing one in every position<sup>3</sup>.

For any norm<sup>4</sup> of the resulting matrix  $\mathcal{R}$ , one can then observe that  $\lim_{N \rightarrow +\infty} \|\mathcal{R}\| = 0$ . 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

<sup>3</sup>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 to a matrix of ones.

<sup>4</sup>For this paper, the spectral norm is used, but any norm would give similar results.

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.

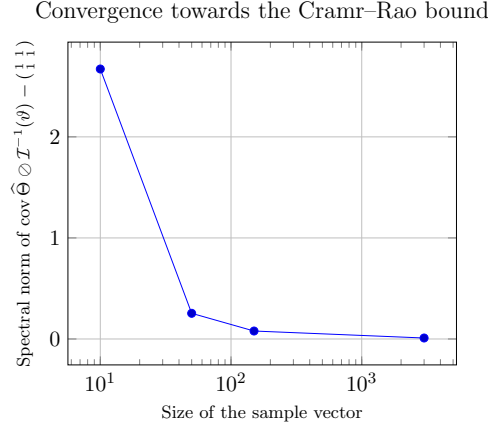


Figure 2. Using an adequately-centered spectral norm, one can visualize to which extent the Cramr–Rao lower bound is reached. The symbol “ $\oslash$ ” denotes Hadamard division [11].

## Part 2. Particle filtering

### 8. Introduction

For the second part of the project, the task is to use noisy observations of the position of a fish in order to compute the parameters of the Gamma distribution governing its speed. With these parameters, a simulation is then completed which models the reaction and movement of the fish to a nearby predator. More precisely, a sequential Monte Carlo method is used to extract the real positions from the noisy measurements. In order to do this, one can play with various parameters, such as the number of fish, the number of particles per fish in the simulation, the time step between observations or the standard deviation of the noise which is added to the position to obtain the observations.

### 9. Estimation of the parameters from noisy measurements

Recall that the speed of the fish follows a Gamma distribution ( $V_i \sim \Gamma(k, s)$ ). Here, one gets noisy observations of the position of a fish at given instants in time. It is rather straightforward to transform a noisy observation of the speed from these positions measurement, knowing the sampling period  $t_s$ . Once we have extracted a noisy approximation of the speed from this set of data, we can repeat what we explained earlier in the document, in Section 5, to estimate the parameters of the distribution.

We begin by defining the formula for the speed more rigorously: from  $x_p(i)$ , the noisy measurement of the position of fish  $p$  at step  $i$ , one can get  $v_p(i)$ , the noisy observation of the speed of that fish at that step, by

$$(9.1) \quad v_p(i) = \frac{\|x_p(i+1) + \mathbf{n}_p(i+1) - (x_p(i) + \mathbf{n}_p(i))\|}{t_s}, \quad \forall p,$$

with  $\mathbf{n}_p(i) \sim \mathcal{N}(0, \sigma_{\text{obs}}^2)$  an additive white noise. Using the estimation techniques of Section 4, one can get estimators of the two parameters of the Gamma distribution of the speed,  $\hat{k}$  and  $\hat{s}$ , based on the observed speeds.

We can see that the zero-mean white noise has no effect on our estimators. Indeed, as shown in (9.1), the white-noise from time  $i + 1$  is "cancelled" with the white-noise from time  $i$ . We hence subtract the bias (which follows the same distribution) of two biased measurements of the position, and it is then "cancelled".

## 10. Sequential Monte Carlo method

We now would like to track the position of the fish and the predator, based on noisy measurements generated by `GenerateObservations`. In this case, one must exclude a linear evolution, hence using a Kalman filter would give bad results; we can however use a sequential Monte Carlo (SMC) method. At each time step, we must generate a large number of samples from the posterior distribution, and use those samples to generate the next state, as explained in [?]. The SMC method does not require us to compute those samples at each time step. The following notations will be used  $f(x_t | y_1, \dots, y_t)$  is the posterior density function at time  $t$ , and  $x_t^i$  is a sample from this posterior density. We can define the sample set:  $\mathcal{S}_t = \{x_t^i\}$ , for  $i = 1, 2, \dots, N_p$ . Our goal is to use the observations in  $\mathcal{S}_t$  and the noisy measurements of the position,  $y_{t+1}^i$  to compute the sample set  $\mathcal{S}_{t+1}$ . After doing so, we can then first create the new sample set, and then correct those values using the noisy measurements.

*Step 1 (Initialization).* When one first calls the function, the initial sample set  $\mathcal{S}_1$  is empty. It is then filled using noisy observations  $y_1^i$ , together with an additive zero-mean white-noise for each particle. This initial guess allows us to start from the only approximation of the position we know (that is, its observed value). To compute the initial orientation of the fish, we use again the noisy measurements from  $y_2^i$  and  $y_1^i$ . Though all the computation, we keep in memory all the sample sets, for each time step, and for each fish and predator. Since we work in a 2-D space area, each sample  $x_t^i$  is a 2-D vector, that is:  $x_t^i = \begin{bmatrix} x & y \end{bmatrix}^T$ .

*Step 2 (Prediction).* To predict the samples at step  $t + 1$ , we can use the posterior density samples at step  $t$ , using the prediction density  $f(x_{t+1} | y_1, \dots, y_t)$ . We can compute the prediction set  $\tilde{\mathcal{S}}_{t+1}$  of samples  $\tilde{x}_{t+1}^i$ . One can show that these last samples are samples from  $f(x_{t+1} | y_1, \dots, y_t)$ . [?]

In our case, we use the `StateUpdate` function to generate the samples at time  $t + 1$ , which gives the state model update of the position.

The basic idea comes from (10.1).

$$(10.1) \quad f(x_{t+1} | y_1, \dots, y_t) = \int_{x_t} f(x_{t+1} | y_1, \dots, y_t) f(x_t | y_1, \dots, y_t) dx_t,$$

Since  $x_t^i$  is a sample from  $f(x_t | y_1, \dots, y_t)$ , we can use a sample from the conditional density  $f(x_{t+1} | x_t^i)$ .

*Step 3 (Correction).* We use the prediction set computed earlier,  $\tilde{\mathcal{S}}_{t+1}$ , to generate posterior samples from the posterior density function. The goal here is to compute weights  $\tilde{w}_{t+1}^i$  for each particle of this set. Those weights allow to give more importance to particles which have a higher probability to be close to the true value of the position at time  $t + 1$ . Hence, the particles which are further from the true value will have a lower weight. These weights are then normalized, which gives  $\tilde{w}_{t+1}^i$ . One can show that if we resample from the set  $\tilde{\mathcal{S}}_{t+1}$  with probabilities given by  $\{\tilde{w}_{t+1}^1, \dots, \tilde{w}_{t+1}^{N_p}\}$ , then the resulting values are samples from the posterior. Generating  $N_p$  resamples, which we call  $x_{t+1}^i$ , gives the sample set

$$(10.2) \quad \mathcal{S}_{t+1} = \{x_{t+1}^1, x_{t+1}^2, \dots, x_{t+1}^{N_p}\},$$

which is the corrected sample set at time  $t + 1$ .



*Step 4 (Termination).* For each time step, we have computed the posterior sample set of the position for each fish and for the predator. However, the true aim of the particle filter is to give a valid approximation of the noiseless value of the position of each fish. One can get these approximations, for time  $t$  and fish  $p$ , by computing the sample mean of the sample set  $\mathcal{S}_t$ , relative to the  $p$ th fish.

## 11. Mean squared error analysis

Using the particle filter of Section 10, one can perform several experiments in order to determine its robustness and dependance on various parameters of the model. In order to quantify this, we use the mean square error, defined as

$$(11.1) \quad E_{\text{MSE}} = \frac{1}{NP} \sum_{p=1}^P \sum_{i=1}^N \|\mathbf{x}_p(i) - \hat{\mathbf{x}}_p(i)\|_2^2.$$

Experiments were conducted for different values of three parameters. The default values were  $w = 20$ ,  $P = 3$ ,  $N = 100$ ,  $N_p = 200$ ,  $t_s = 0.1$  and  $\sigma_{\text{obs}} = 0.2$ . All experiments were repeated  $M = 100$  times.

First, the effect of varying the number of particles per fish was studied. Simulations were run with  $N_p = 10, 20, 50, 100, 200, 500, 1000$ . The results are shown on Figure 3. The figure shows

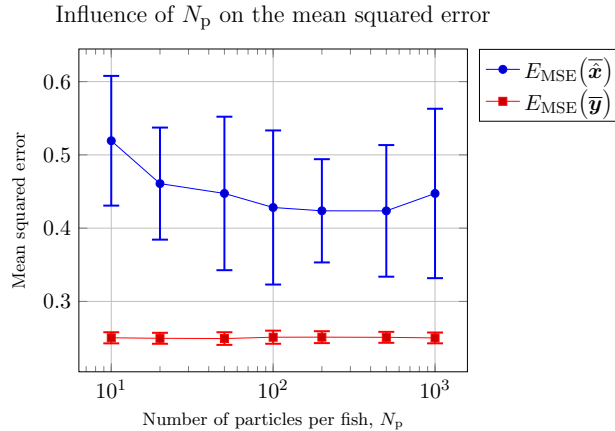


Figure 3. Mean squared error resulting from the filtered position estimates, for varying values of  $N_p$ .

that the mean squared error barely depends on the value of  $N_p$ , if the parameter is sufficiently big ( $N_p \geq 10$  seems to work). We can also observe that the mean squared error resulting from simply passing the observed positions as estimates is lower than the error with the particle filter.

Next, we studied the influence of the sampling period  $t_s$ . Simulations were run with  $t_s = 0.05, 0.1, 0.2, 0.5, 1, 2, 5$ . The results are shown on Figure 4. In line with physical intuition, the larger the sampling period, the less accurate the results become. A sampling period of  $t_s = 0.2$  s is the maximum acceptable value, with an average error around 1. Again, we observe that simply passing the positional observations as estimated positions yields better results than using the conventional filter.

Finally, the last parameter whose influence was examined is the standard deviation on the observed values,  $\sigma_{\text{obs}}$ . Simulations were run with  $\sigma_{\text{obs}} = 0.01, 0.1, 0.2, 0.5, 1, 2, 5$ . The results are shown on Figure 5.

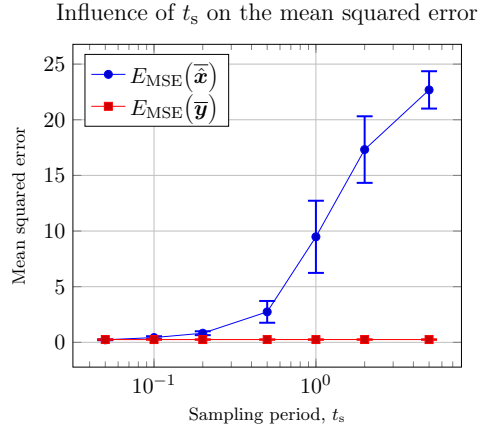


Figure 4. Mean squared error resulting from the filtered position estimates, for varying values of  $t_s$ .

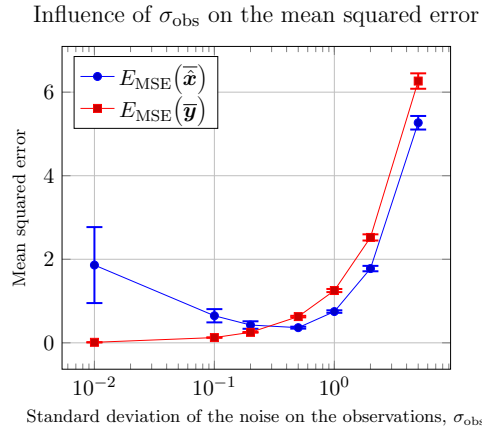


Figure 5. Mean squared error resulting from the filtered position estimates, for varying values of  $\sigma_{\text{obs}}$ .

## 12. Conclusion

In Section 2, we looked at maximum likelihood estimation for the scale parameter of a Gamma distribution,  $s$ . We found the result to be  $\hat{s}_{\text{ML}} = \frac{\bar{V}}{k}$  in (2.8). We proved some asymptotic properties of this estimator in Section 3, showing that it is asymptotically unbiased, efficient, best asymptotically normal and consistent. After establishing this, in Section 4, we looked into estimating both parameters of the Gamma distribution jointly, using our previous estimator. This led us to find that  $\hat{k}_{\text{ML}}$  is the solution of (4.5). This estimator has no closed form, hence in Section 5 we turned to numerical methods to estimate it; comparing the maximum likelihood estimators with the method of moments estimators showed that both are unbiased (for both parameters), with the former generally having lower variance, while the latter can be useful as an initial guess for the maximum likelihood estimator. Section 6, together with Appendix C, details an analytical derivation of the bound given by the Cramér–Rao inequality of Theorem A.1. Finally, in Section 7, empirical evidence of the efficiency of the maximum likelihood estimator was found.

## Appendix A. Definitions of properties

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

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

**THEOREM A.1** (Cramr–Rao inequality). *If  $Z = (Z_1, \dots, Z_N)^T$  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 Cramr–Rao inequality*

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

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

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

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

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

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

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

for some minimal positive definite matrix  $\Sigma$ .

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

$$(A.6) \quad \text{plim}_{N \rightarrow +\infty} \widehat{\Theta}_N(Z) = \vartheta.$$

Theorem 9.1 of [6] proves that this is equivalent to the statement that the MSE of the estimator converges to zero as  $N$  goes to infinity.

## Appendix B. Omitted proofs

The following is a proof of asymptotic normality for the estimator in (2.8), using the Lindeberg–Lvy formulation of the central limit theorem.

*Proof.* We want to prove that  $\frac{1}{\sqrt{N}} \ell'(\vartheta_0) \sim \mathcal{N}(0, \mathcal{I}(\vartheta_0))$ . We simplify notation by writing  $f(v_i)$  instead of  $f_{V_i}(v_i; k, \vartheta)$  and using Euler’s notation for partial derivatives [14]. First, we show that

the expected value of  $\frac{1}{\sqrt{N}}\ell'(\vartheta_0)$  is zero.

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

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

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

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

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

$$(B.6) \quad = 0.$$

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

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

$$(B.8) \quad = \int_{-\infty}^{+\infty} \partial_{\vartheta_0} \ln f(v_i) \partial_{\vartheta_0} f(v_i) dv_i.$$

Using the product rule, we can then find

$$(B.9) \quad = - \int_{-\infty}^{+\infty} \partial_{\vartheta_0 \vartheta_0} \ln f(v_i) f(v_i) dv_i \\ + \int_{-\infty}^{+\infty} \partial_{\vartheta_0} \left( \partial_{\vartheta_0} \ln f(v_i) f(v_i) \right) dv_i$$

$$(B.10) \quad = -\mathbb{E} [\partial_{\vartheta_0 \vartheta_0} \ln f(v_i)] + \partial_{\vartheta_0} \int_{-\infty}^{\infty} \frac{\partial_{\vartheta_0} f(v_i)}{f(v_i)} f(v_i) dv_i$$

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

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

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

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

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

### 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  $\ell(v; k, s)$

instead of  $\ell(v_1, \dots, v_N; k, s)$ :

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

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

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

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

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

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

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

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

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

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

## References

- [1] I. DUNNING, J. HUCHETTE, and M. LUBIN, JuMP: A Modeling Language for Mathematical Optimization, *SIAM Review* **59** (2017), 295–320. <http://dx.doi.org/10.1137/15M1020575>.
- [2] ERIC W. WEISSTEIN, Digamma Function, From MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/DigammaFunction.html>, [Online; accessed 9-April-2019].
- [3] ERIC W. WEISSTEIN, Trigamma Function, From MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/TrigammaFunction.html>, [Online; accessed 10-April-2019].
- [4] A. HUTH and C. WISSEL, The Simulation of the Movement of Fish Schools, *Journal of Theoretical Biology* **156** (1992), 365–385.
- [5] THOMAS P. MINKA, Estimating a Gamma distribution, <https://tminka.github.io/papers/minka-gamma.pdf>, 2002, [Online; accessed 6-April-2019].
- [6] D. D. WACKERLY, W. MENDENHALL III, and R. L. SCHEAFFER, *Mathematical Statistics with Applications*, 7 ed., Brooks/Cole, Cengage Learning, 20 Davis Drive, Belmont, CA 94002, USA, 2008.
- [7] WIKIPEDIA CONTRIBUTORS, Sample mean and covariance — Wikipedia, the free encyclopedia, [https://en.wikipedia.org/wiki/Sample\\_mean\\_and\\_covariance#Sample\\_covariance](https://en.wikipedia.org/wiki/Sample_mean_and_covariance#Sample_covariance), 2018, [Online; accessed 6-April-2019].
- [8] WIKIPEDIA CONTRIBUTORS, Central limit theorem — Wikipedia, The Free Encyclopedia, [https://en.wikipedia.org/wiki/Central\\_limit\\_theorem](https://en.wikipedia.org/wiki/Central_limit_theorem), 2019, [Online; accessed 12-April-2019].
- [9] WIKIPEDIA CONTRIBUTORS, Characteristic function (probability theory) — Wikipedia, The Free Encyclopedia, [https://en.wikipedia.org/wiki/Characteristic\\_function\\_\(probability\\_theory\)](https://en.wikipedia.org/wiki/Characteristic_function_(probability_theory)), 2019, [Online; accessed 12-April-2019].
- [10] WIKIPEDIA CONTRIBUTORS, Gamma distribution — Wikipedia, The Free Encyclopedia, [https://en.wikipedia.org/wiki/Gamma\\_distribution](https://en.wikipedia.org/wiki/Gamma_distribution), 2019, [Online; accessed 11-April-2019].

- [11] WIKIPEDIA CONTRIBUTORS, Hadamard product (matrices) — Wikipedia, The Free Encyclopedia, [https://en.wikipedia.org/wiki/Hadamard\\_product\\_\(matrices\)#Analogous\\_operations](https://en.wikipedia.org/wiki/Hadamard_product_(matrices)#Analogous_operations), 2019, [Online; accessed 17-April-2019].
- [12] WIKIPEDIA CONTRIBUTORS, Matrix norm — Wikipedia, the free encyclopedia, [https://en.wikipedia.org/wiki/Matrix\\_norm#Spectral\\_norm](https://en.wikipedia.org/wiki/Matrix_norm#Spectral_norm), 2019, [Online; accessed 17-April-2019].
- [13] WIKIPEDIA CONTRIBUTORS, Maximum likelihood estimation — Wikipedia, The Free Encyclopedia, [https://en.wikipedia.org/wiki/Maximum\\_likelihood\\_estimation](https://en.wikipedia.org/wiki/Maximum_likelihood_estimation), 2019, [Online; accessed 8-April-2019].
- [14] WIKIPEDIA CONTRIBUTORS, Notation for differentiation — Wikipedia, the free encyclopedia, [https://en.wikipedia.org/wiki/Notation\\_for\\_differentiation#Euler's\\_notation](https://en.wikipedia.org/wiki/Notation_for_differentiation#Euler's_notation), 2019, [Online; accessed 12-April-2019].

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