



Master de Statistique 1^{ère} année

01 Juin 2012 au 31 Juillet 2012

Rapport de stage

Analyse de mouvements de fourragements
chez le guillemot de Brünnich
(*Uria lomvia*)

Stagiaire : Mohamed AKODAD

Maître de stage : M. Cédric SUEUR

REMERCIEMENTS

Je tiens tout d'abord à remercier Mr Cédric SUEUR, maître de conférences à l'université de Strasbourg, ainsi que Mr Nicolas POULIN, ingénieur de recherche, pour m'avoir encadré durant ces deux mois de stage, pour le temps qu'ils m'ont consacré tout au long de cette période et pour avoir répondu à mes questions.

Je remercie également, Mme Odile PETIT, responsable de l'équipe, pour m'avoir accueilli, pour ses conseils et recommandations.

Je tiens à remercier également, Mme Valérie DUFOUR et Céline BRET, pour leur accueil sympathique et leur aide tout au long de ces deux mois.

Un grand merci également à toutes les personnes rencontrées au sein du DEPE durant ces deux mois de stage pour leur accueil et leur bonne humeur, en particulier à Yan ROPERT COUDERT et Akiko KATO.


Table des matières

1) Introduction	4
1.1) Présentation du laboratoire	4
1.2) Données et présentation du sujet	5
2) Les lois de probabilités	7
2.1) La loi exponentielle	7
2.2) La loi de Weibull	7
2.3) Loi de type puissance	8
3) Estimation par maximum de vraisemblance.....	9
3.1) Modèles statistiques	9
3.2) La vraisemblance.....	9
3.3) Estimation par maximum de vraisemblance.....	9
3.3.1) Calcul de l'estimation par maximum de vraisemblance pour la loi exponentielle.....	10
3.3.2) Résultats :	11
3.3.3) Calcul de l'estimation par maximum de vraisemblance pour la loi de weibull.....	12
3.3.4) Résultats :	13
4) Test Statistique et QQ-plot	14
4.1) Test statistique.....	14
4.2) Test de Kolmogorov-Smirnov	15
4.2.1) Résultats :	15
4.3) QQ-Plot	16
4.3.1) Résultats :	17
5) DISCUSSION	21
6) Bibliographie	30

1) Introduction

1.1) Présentation du laboratoire

J'ai réalisé mon stage du 01 Juin 2012 au 31 Juillet 2012, au Département d'Ecologie, Physiologie et éthologie (DEPE) dirigé par Mme Odile Petit et qui est un des trois départements de l'Institut Pluridisciplinaire Hubert Curien (IPHC).

 Département d'Ecologie, Physiologie et Ethologie - DEPE Responsable : Odile PETIT (DR) Adjoint : Jean Patrice ROBIN (CR)		Mars 2012		
MOYENS TECHNIQUES Animalerie Caroline Habold (CR) Aurélie Hranitsky (AUT) Biochimie Mathilde Arrivé (AI) Biologie Moléculaire Sandrine Zahn (IE) Ecologie Moléculaire Héliène Gachot (IR) Microscopie à balayage Nicole Liewig (CR) Spectrométrie de masse isotope Alexandre Zahariev (AI) Isabelle Chery (TCH)		Equipes scientifiques		
ADMINISTRATION Gestion financière et comptable Jacqueline Brencklé (TCH) Claudine Gallone (AI) Ressources humaines Martine Schneider (TCH) Missions et secrétariat		Ecophysiologie évolutive François CRISCUOLO (CR) Fabielle Angel (CR) Stéphane Blanc (DR) Caroline Habold (CR) Nicole Liewig (CR) Sylvie Massemin (MCF) Hugues Oudart (MCF) Jean Patrice Robin (CR) Mathilde Arrivé (AI) Isabelle Chery (TCH) Alexandre Zahariev (AI) Sandrine Zahn (IE)	Ecologie fonctionnelle Yan ROPERT-COUDERT (CR) André Ancel (CR) Jean Yves Georges (CR) Yves Handrich (CR) Yvon Le Maho (DR) Thierry Raclot (CR) Magali Beauguey (AI) Damien Chevallier (IR) Héliène Gachot (IR)	Ethologie évolutive Odile PETIT (DR) Valérie Dufour (CR) Ronald Noë (PR) Jean Jacques Roeder (DR) Cédric Sueur (CR) Bernard Thierry (DR) Pierre Uhlrich (TCH)
SOUTIEN A LA RECHERCHE Documentation/Web Brigitte Gaillard (AI) Informatique Yannick Bertho (AI)		Equipe technique : Métrologie et Instrumentation en Biologie et Environnement (MIBE) Francis Crenner (IR) Nicolas Chatelain (AI) Akiko Kato (IR) Thê Duc Le (AI) Christian Weber (IE) Julien Courtecuisse (IR) Robin Laesser (IE) Marc Richer (IE)		Stations de terrain : * Alsace Station d'Etude de la Biodiversité (SEB) Bernard Thierry (DR) Pierre Uhlrich (TCH) * Guyane Yvon Le Maho (DR) Damien Chevallier (IR)
SECURITE ACMO Alexandre Zahariev (AI) PCR Mathilde Arrivé (AI) Qualité Héliène Gachot (IR)				

Ce département possède trois axes de recherche :

- Le premier étudie comment l'animal répond aux changements environnementaux (variations climatiques et/ou abondance des ressources alimentaires) ou anticipe ces changements.
- Le deuxième caractérise les mécanismes adaptatifs, comportementaux ou physiologiques, avec leurs éventuelles implications biomédicales ou biotechnologiques.
- Le troisième utilise les animaux comme indicateurs des conséquences des changements climatiques sur les ressources alimentaires.

1.2) Données et présentation du sujet

La collecte des données a eu lieu en Juillet 2011, à Coats Island, Nunavut (CANADA), par l'équipe de Yan Ropert Coudert à l'IPHC, sur quatorze guillemots de Brünnich (*Uria lomvia*). Le guillemot est un oiseau marin plongeur de la famille des alcidés (figure 1). Comme la plupart des oiseaux marins le guillemot est un fourrageur à place centrale (« central place forager » en Anglais) qui niche sur des falaises pendant le court été boréal, ce qui impose une contrainte temporelle sur la réalisation de la reproduction. Les parents réalisent des trajets en mer soit pour se nourrir lors de la phase d'incubation de l'œuf, soit pour approvisionner le poussin lorsque l'œuf a éclot. Durant les trajets en mer, ils alternent des phases de vol battu et de vol plané pour se rendre sur les zones riches en proies (diverses espèces de poissons), avec des phases de plongées plus ou moins profondes durant lesquelles ils poursuivent leurs proies sous l'eau. La collecte des données s'est faite grâce à des bio-loggers.

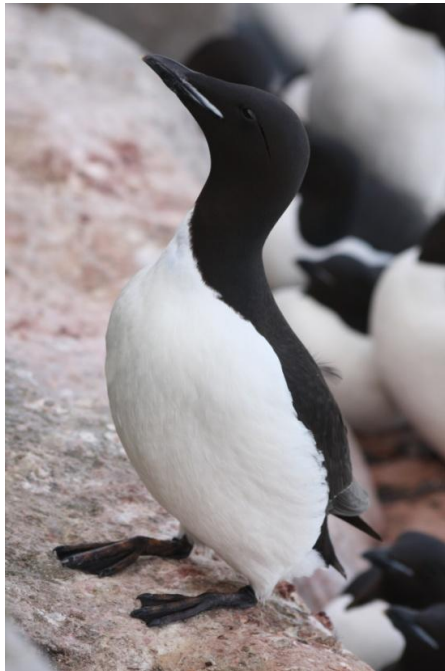


Figure 1 : Individu adulte de l'espèce guillemot de Brünnich

Un bio-logger est un dispositif qui se fixe sur l'animal et qui permet d'enregistrer les données biologiques d'un être vivant, telles que les déplacements, la vitesse d'accélération, et bien d'autres paramètres biologiques (Ropert-Coudert & Wilson 2005). Le guillemot peut être équipé soit d'un GPS uniquement, soit d'un GPS en combinaison avec d'autres instruments de mesures, comme par exemple un enregistreur de la température et de la pression hydrostatique pour les oiseaux plongeurs.

Dans notre étude, je me suis intéressé aux données obtenues par les GPS, notamment la distance parcourue par les guillemots, distance qui était enregistrée toutes les deux minutes. Le calcul de la distance s'est fait par l'intermédiaire de la latitude et la longitude relevé par le GPS. De plus, je distinguerai deux types de données, les données brutes, issues du calcul direct de la distance et des données arrondies, données issues des données brutes dont nous avons pris l'arrondi à l'entier le plus proche.

Plusieurs articles ont été écrits quant à l'étude des mouvements de fourragements chez les animaux (e.g. Bartumeus et al. 2012, Marell et al. 2002, Reynolds 2012, Sueur et al. 2011, Sueur 2011) afin de mieux comprendre leurs déplacements dans leurs écosystèmes. En effet, dans ces articles, ils étudient la loi de distribution qui décrit le mieux les déplacements et en déduisent si le mouvement est de type brownien ou Lévy. Le mouvement brownien correspond à une recherche aléatoire tandis que le vol de Lévy correspond à une recherche dirigée et optimale de l'animal. Si les données suivent une loi exponentielle alors le mouvement est brownien, en revanche, si elles suivent une loi de type puissance alors le mouvement correspond à un vol de Lévy. En effet, le vol de Lévy produit des trajectoires avec de plusieurs petits déplacements et temporairement des déplacements longs, tandis que le vol brownien produit des déplacements constants.

C'est pour cela que, nous étudierons les déplacements des guillemots dans un premier temps, nous allons estimer par maximum de vraisemblance les paramètres d'une loi de Weibull et le paramètre d'une loi exponentielle, puis, nous testerons si ces données correspondent à une loi de Weibull car elle possède une queue de distribution plus longue que la loi de type puissance, mais nous testerons également si ces données correspondent à une loi exponentielle car elle représente une loi particulière de la loi de Weibull ($\beta = 1$) et possède une queue de distribution plus petite.

2) Les lois de probabilités

2.1) La loi exponentielle

Définition :

Soit $\lambda > 0$. Une variable aléatoire X à valeurs dans $[0; +\infty[$ suit la **loi exponentielle de paramètre λ** , si X est une variable continue et admet pour densité de probabilité la fonction suivante :

$$f_X(t) = \begin{cases} \lambda e^{-\lambda t} & \text{pour } t \geq 0 \\ 0 & \text{pour } t < 0 \end{cases}$$

Proposition :

La fonction de répartition d'une loi exponentielle est égale à :

$$F(x) = \begin{cases} 1 - \exp(-\lambda x) & \text{si } x \geq 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

2.2) La loi de Weibull

Définition :

Soit $\beta > 0$ et $\eta > 0$. Une variable aléatoire X à valeurs dans $[0; +\infty[$ suit la **loi de Weibull de paramètres β et η** , si X est une variable continue et admet pour densité de probabilité la fonction suivante :

$$f_{\beta,\eta}(t) = \begin{cases} \frac{\beta}{\eta^\beta} t^{\beta-1} \exp\left(-\left(\frac{t}{\eta}\right)^\beta\right) & \text{pour } t \geq 0 \\ 0 & \text{pour } t < 0 \end{cases}$$

Proposition :

La fonction de répartition d'une loi de Weibull est égale à :

$$F(x) = \begin{cases} 1 - \exp\left(-\left(\frac{x}{\eta}\right)^\beta\right) & \text{si } x \geq 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

2.3) Loi de type puissance

Ce type de loi est utilisé par Bartumeus et al. (2012), Marell et al. (2002), Sueur et al. (2011), et Sueur, (2011), mais j'ai trouvé très peu de bibliographie en mathématiques évoquant ce type de loi, peut être veulent ils parler de loi de type puissance, comme on parle de lois appartenant à la famille exponentielle.

Définition :

Soit $\alpha > 1$. Une variable aléatoire X à valeurs dans $[x_{min}; +\infty[$ suit une **loi de type puissance de paramètres α et x_{min}** , si X admet pour densité de probabilité la fonction suivante :

$$f_X(t) = \begin{cases} \frac{\alpha-1}{x_{min}} \left(\frac{t}{x_{min}}\right)^{-\alpha} & \text{pour } t \geq x_{min} \\ 0 & \text{pour } t < x_{min} \end{cases}$$

La fonction de répartition d'une loi de type puissance est égale à :

$$F(x) = \begin{cases} 1 - \left(\frac{x}{x_{min}}\right)^{-\alpha+1} & \text{si } x \geq x_{min} \\ 0 & \text{si } x < x_{min} \end{cases}$$

3) Estimation par maximum de vraisemblance

3.1) Modèles statistiques

L'introduction de modèles statistiques est un moyen de prendre en compte l'information à priori. On va supposer que la loi inconnue \mathbf{P}_X appartient à un ensemble de lois déterminées à l'avance, celles présentées ci-dessus.

Définition :

Soit X une variable aléatoire dans (X, F) , où F est la tribu borélienne de X .

On appelle **modèle statistique** pour X tout triplet $(X, F, \{\mathbf{P}_\theta\}_{\theta \in \Theta})$, où $\{\mathbf{P}_\theta\}_{\theta \in \Theta}$ est une collection de probabilités sur (X, F) indexée par un ensemble d'indices quelconque Θ tel que $\mathbf{P}_X \in \{\mathbf{P}_\theta\}_{\theta \in \Theta}$.

Un **modèle statistique est dominé** s'il existe une mesure σ -finie μ sur (X, F) telle que :

Pour tout $F \in F$ $\mu(F) = 0 \Rightarrow$ pour tout $\theta \in \Theta$ $\mathbf{P}_\theta(F) = 0$

3.2) La vraisemblance

D'après le théorème de Radon-Nikodym, si $(X, F, \{\mathbf{P}_\theta\}_{\theta \in \Theta})$ est dominé par une mesure μ , il existe pour tout $\theta \in \Theta$ une densité $f_\theta : X \rightarrow \mathbf{R}^+$

On peut définir la vraisemblance du modèle $(X, F, \{\mathbf{P}_\theta\}_{\theta \in \Theta})$ de la façon suivante :

$$L : X \times \Theta \rightarrow \mathbf{R}^+ \\ (x, \theta) \rightarrow f_\theta(x)$$

3.3) Estimation par maximum de vraisemblance

Si f est la densité d'une certaine loi de probabilité de paramètre θ inconnu, et si nous avons n observations x_1, \dots, x_n l'estimateur du maximum de vraisemblance de θ est donné par :

$$\hat{\theta} = \operatorname{argmax}_{\theta \in \mathbf{R}} \prod_{i=1}^n f(x_i; \theta)$$

Ou encore en passant au logarithme :

$$\hat{\theta} = \operatorname{argmax}_{\theta \in \mathbf{R}} \sum_{i=1}^n \log(f(x_i; \theta))$$

Dans notre cas, les données sont considérées comme indépendantes, comme il a été décrit dans de précédents articles (Bartumeus et al. 2012, Marell et al. 2002, Reynolds, 2012, Sueur et al. 2011 et Sueur, 2011).

3.3.1) Calcul de l'estimation par maximum de vraisemblance pour la loi exponentielle

La vraisemblance est un produit de valeurs de la densité et elle vaut :

$$L(x_1, \dots, x_n, \lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum x_i}$$

Son logarithme est :

$$\log(L(x_1, \dots, x_n, \lambda)) = n \log(\lambda) - \lambda \sum x_i$$

La dérivée par rapport à λ est :

$$\frac{\partial \log(L(x_1, \dots, x_n, \lambda))}{\partial \lambda} = \frac{n}{\lambda} - \sum x_i$$

L'estimateur de maximum de vraisemblance vaut :

$$\hat{\lambda} = \frac{n}{\sum x_i}$$

En effet la dérivée seconde est négative est vaut :

$$\frac{-n}{\lambda^2}$$

3.3.2) Résultats :

Tableau 1 : résultats de l'estimation des données brutes

	$\hat{\lambda}$
guillemot 1	0.008837029
guillemot 2	0.007126994
guillemot 3	0.005439038
guillemot 4	0.01385652
guillemot 5	0.008462279
guillemot 6	0.00549153
guillemot 7	0.006618427
guillemot 8	0.005187178
guillemot 9	0.005452742
guillemot 10	0.003699699
guillemot 11	0.00519025
guillemot 12	0.006919538
guillemot 13	0.008773009
guillemot 14	0.007747013

Tableau 2 : résultats de l'estimation des données arrondies:

	$\hat{\lambda}$
guillemot 1	0.002010774
guillemot 2	0.002302519
guillemot 3	0.001892101
guillemot 4	0.005054313
guillemot 5	0.001854129
guillemot 6	0.003226694
guillemot 7	0.001536615
guillemot 8	0.003147999
guillemot 9	0.001254653
guillemot 10	0.001854294
guillemot 11	0.003712504
guillemot 12	0.001515957
guillemot 13	0.004738214
guillemot 14	0.004657582

3.3.3) Calcul de l'estimation par maximum de vraisemblance pour la loi de weibull

La vraisemblance est un produit de valeurs de la densité et elle vaut :

$$\begin{aligned} L(x_1, \dots, x_n, \eta, \beta) &= \prod_{i=1}^n \frac{\beta}{\eta^\beta} x_i^{\beta-1} \exp\left(-\left(\frac{x_i}{\eta}\right)^\beta\right) \\ &= \left(\frac{\beta}{\eta^\beta}\right)^n \prod_{i=1}^n x_i^{\beta-1} \exp\left(-\sum_{i=1}^n \left(\frac{x_i}{\eta}\right)^\beta\right) \end{aligned}$$

Son logarithme est :

$$\log(L(x_1, \dots, x_n, \eta, \beta)) = n \log(\beta) - n\beta \log(\eta) + (k-1) \sum_{i=1}^n \log(x_i) - \sum_{i=1}^n \left(\frac{x_i}{\eta}\right)^\beta$$

La dérivée par rapport à η est :

$$\frac{\partial \log(L(x_1, \dots, x_n, \eta, \beta))}{\partial \eta} = \frac{-n\beta}{\eta} + \frac{\beta}{\eta^{\beta+1}} \sum_{i=1}^n x_i^\beta$$

La dérivée par rapport à β est :

$$\frac{\partial \log(L(x_1, \dots, x_n, \eta, \beta))}{\partial \beta} = \frac{-n}{\beta} - n \log(\eta) + \sum_{i=1}^n \log(x_i) - \sum_{i=1}^n \log(x_i) \left(\frac{x_i}{\eta}\right)^\beta$$

Il n'existe pas d'expression analytique pour trouver $\hat{\eta}$ et $\hat{\beta}$, comme pour le calcul fait ci-dessus pour l'estimateur du maximum de vraisemblance pour une loi exponentielle.

C'est pour cela, que nous allons utiliser des fonctions du logiciel R, afin d'en approximer les solutions.

Nous allons utiliser les fonctions suivantes :

-`fitdistr()` qui est incluse dans le package MASS

-`mle()` qui est incluse dans le package stats4

La fonction `fitdistr()` permet de calculer, l'estimation par maximum de vraisemblance de quasiment toutes les lois, en particulier la loi de Weibull, en voici un extrait du code :

```
x1<-fitdistr(Step.length, "weibull")
```

où `Step.length` représente les données et “weibull”, la loi dont on veut estimer les paramètres.

Ensuite, la fonction `mle()` permet d'optimiser l'estimation des paramètres, en calculant le maximum de la vraisemblance de manière itérative, mais cette fonction a besoin de valeurs de départ, on prendra celles données par la fonction `fitdistr()` car comme on l'a constaté, il n'existe pas d'expression analytique afin de trouver $\hat{\eta}$ et $\hat{\beta}$, en voici un extrait du code :

```
par<-c(0.61488647,61.46620451)
```

```
vrais<-function (a,b){
```

```
n=length(Step.length)
```

```
logvrais=n*log(a)-n*a*log(b)+(a-1)*sum(log(Step.length))-sum((Step.length/b)^a)
```

```
return(-logvrais)}
```

```
est<-mle(minuslog=vrais, start=list(a=0.61488647,b=61.46620451))
```

```
summary(est)
```

où « `par` » représente l'estimation par la fonction `fitdistr()`,

`vrais()`, représente le logarithme de la vraisemblance d'une loi de Weibull,

`summary(est)`, renvoie l'estimation par maximum de vraisemblance et $-\log(L)$ dont nous allons nous servir afin de calculer l'AIC et le BIC.

3.3.4) Résultats :

Tableau 3 : résultats de l'estimation des données brutes

	$\hat{\beta}$	$\hat{\eta}$
guillemot 1	0.6148905	61.4662048
guillemot 2	0.5513691	68.1633476
guillemot 3	0.6140508	108.6406267
guillemot 4	0.9843881	71.5002051
guillemot 5	0.6357845	67.6314294
guillemot 6	0.7368288	137.3954858
guillemot 7	0.6277163	84.8884706
guillemot 8	0.7750359	155.707396
guillemot 9	0.5082454	66.5909475
guillemot 10	0.5825709	142.6095753
guillemot 11	0.8107832	164.832088
guillemot 12	0.6116142	77.0117624
guillemot 13	0.8635539	101.928885
guillemot 14	0.7386771	97.5919065

Tableau 4 : résultats de l'estimation des données arrondies

	$\hat{\beta}$	$\hat{\eta}$
guillemot 1	0.6244416	333.2814369
guillemot 2	0.6994197	337.7280902
guillemot 3	0.7262906	427.7491936
guillemot 4	0.8629034	181.3391628
guillemot 5	0.621403	361.811284
guillemot 6	0.6712134	216.1714808
guillemot 7	0.5808569	391.5476366
guillemot 8	0.7028042	235.2334209
guillemot 9	0.5731174	501.5219949
guillemot 10	0.6213688	298.4377274
guillemot 11	0.7349968	211.4136884
guillemot 12	0.5752378	403.9340379
guillemot 13	0.7199844	157.4868557
guillemot 14	0.6823956	151.874117

4) Test Statistique et QQ-plot

4.1) Test statistique

Définition :

On considère un modèle statistique suivant : $(\mathcal{X}, \mathcal{F}, \{\mathbf{P}_\theta\}_{\theta \in \Theta})$

Un **test** de H_0 contre H_1 est **une statistique** ϕ de $(\mathcal{X}, \mathcal{F})$ dans $\{0, 1\}$ à laquelle on associe la décision : $\begin{cases} \text{si } \phi(\underline{x}) = 0 \text{ on décide } H_0 \text{ vraie} \\ \text{sinon on décide } H_1 \text{ vraie} \end{cases}$

Un test est un mécanisme qui permet de statuer entre les deux hypothèses à la vue des résultats d'un échantillon. La décision du test aboutira à choisir H_0 ou H_1 .

Tableau 5 : On peut envisager quatre cas de figure

		on décide H_0 vraie	on décide H_1 vraie
Réalité	H_0 vraie	bonne décision	erreur de première espèce
	H_1 vraie	erreur de seconde espèce	bonne décision

L'erreur de première espèce est noté α , c'est la probabilité de décider H_1 alors que H_0 est vraie.

L'erreur de seconde espèce est noté β , c'est la probabilité de décider H_0 alors que H_1 est vraie.

Pour les tests réalisés, nous prendrons un risque de première espèce $\alpha=5\%$

Soit ϕ une statistique de test de réalisation ϕ_{obs} , la p-valeur est un réel compris entre 0 et 1 tel que $p = P_{H_0} (|\phi| > \phi_{obs})$.

Si $p > 0.05$, le test n'est pas significatif et nous conservons H_0 .

Si $p < 0.05$, le test est significatif, nous rejetons H_0 et nous conservons H_1 .

4.2) Test de Kolmogorov-Smirnov

Le test de Kolmogorov-Smirnov permet de tester si un échantillon suit une loi donnée mais aussi de tester si deux échantillons suivent la même loi, néanmoins, il ne s'applique qu'à des distributions continues.

Le test de Kolmogorov-Smirnov est un test d'adéquation à une loi théorique dont la statistique de test pour une fonction de répartition $F(x)$ est :

$$D = \sup |F_n(x) - F(x)|$$

où F_n est la fonction de répartition empirique pour n observations indépendantes et identiquement distribuées. Dans notre étude on s'est intéressé à deux lois en particuliers, la loi exponentielle et la loi de Weibull.

4.2.1) Résultats :

Tableau 6 : résultats du test de Kolmogorov-Smirnov des données brutes

	p-valeur	
	loi de weibull	loi exponentielle
guillemot 1	<2.2e-16	<2.2e-16
guillemot 2	2.51E-14	<2.2e-16
guillemot 3	<2.2e-16	<2.2e-16
guillemot 4	<2.2e-16	<2.2e-16
guillemot 5	<2.2e-16	<2.2e-16
guillemot 6	2.65E-14	<2.2e-16
guillemot 7	<2.2e-16	<2.2e-16
guillemot 8	2.22E-15	<2.2e-16
guillemot 9	<2.2e-16	<2.2e-16
guillemot 10	<2.2e-16	<2.2e-16
guillemot 11	1.67E-11	<2.2e-16
guillemot 12	<2.2e-16	<2.2e-16
guillemot 13	2.12E-11	2.22E-16
guillemot 14	9.95E-10	<2.2e-16

Tableau 7 : résultats du test de Kolmogorov-Smirnov des données arrondies

	p-valeur	
	loi de weibull	loi exponentielle
guillemot 1	0.003415	<2.2e-16
guillemot 2	0.008995	3.98E-12
guillemot 3	0.002098	1.01E-13
guillemot 4	0.03248	0.0003073
guillemot 5	0.00004549	<2.2e-16
guillemot 6	0.001642	3.04E-13
guillemot 7	1.84E-06	<2.2e-16
guillemot 8	3.16E-05	1.72E-14
guillemot 9	2.18E-06	<2.2e-16
guillemot 10	8.86E-07	<2.2e-16
guillemot 11	0.0005055	1.74E-10
guillemot 12	3.87E-06	<2.2e-16
guillemot 13	0.0005639	3.95E-11
guillemot 14	0.0004183	5.46E-12

4.3) QQ-Plot

Un diagramme quantile-quantile, « Q-Q plot », permet une évaluation graphique de la concordance entre une distribution observée et un modèle théorique.

On s'intéresse toujours à la loi exponentielle et à la loi de Weibull, les résultats qui suivent montrent les diagrammes quantile-quantile pour seulement le premier guillemot, les autres tracés ont été effectués mais ne sont pas mis dans le rapport car il y aurait beaucoup de tracés.

4.3.1) Résultats :

Les diagrammes quantile-quantile des données brutes sont les suivants :

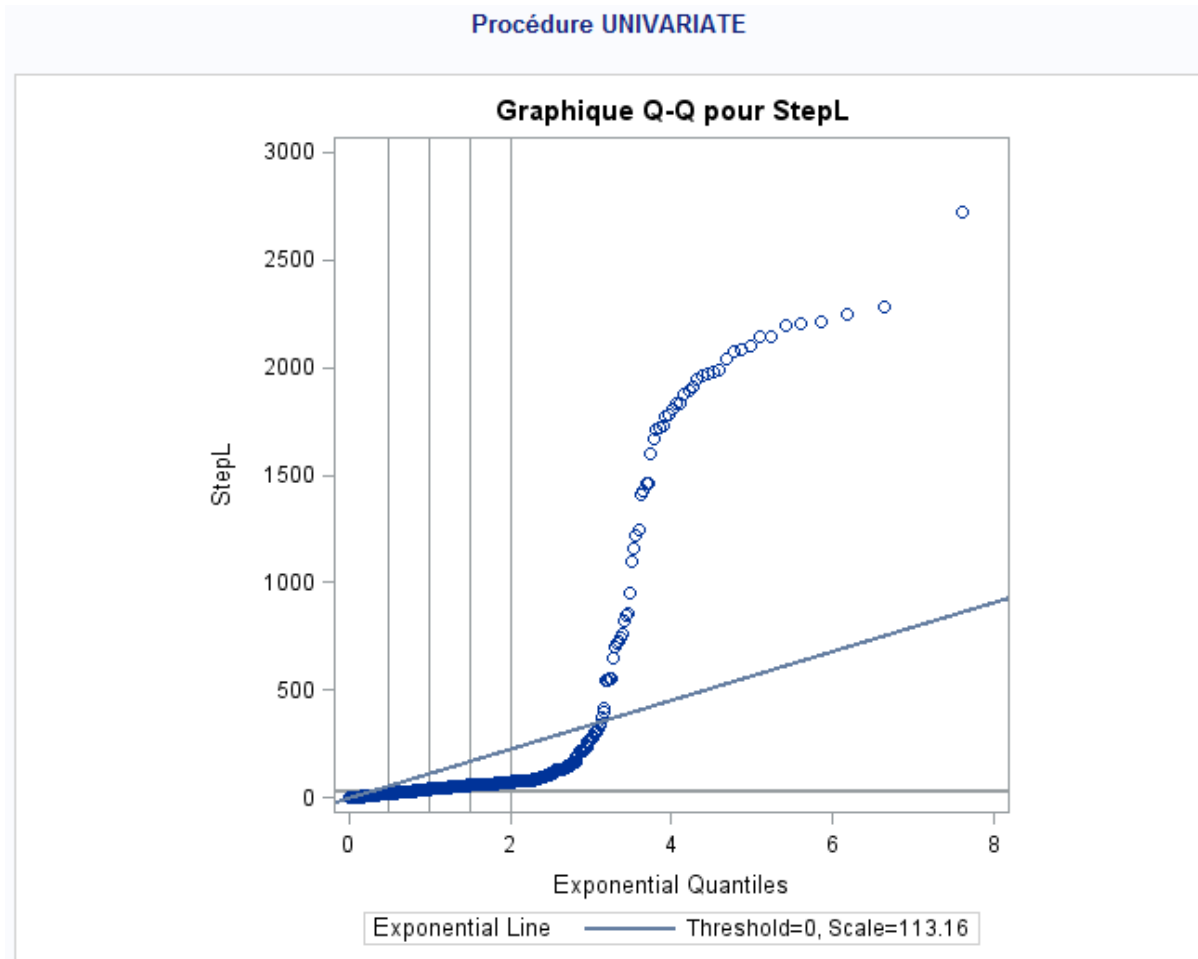


Figure 1 : diagramme quantile-quantile de la concordance des données brutes à une loi exponentielle

Procédure UNIVARIATE

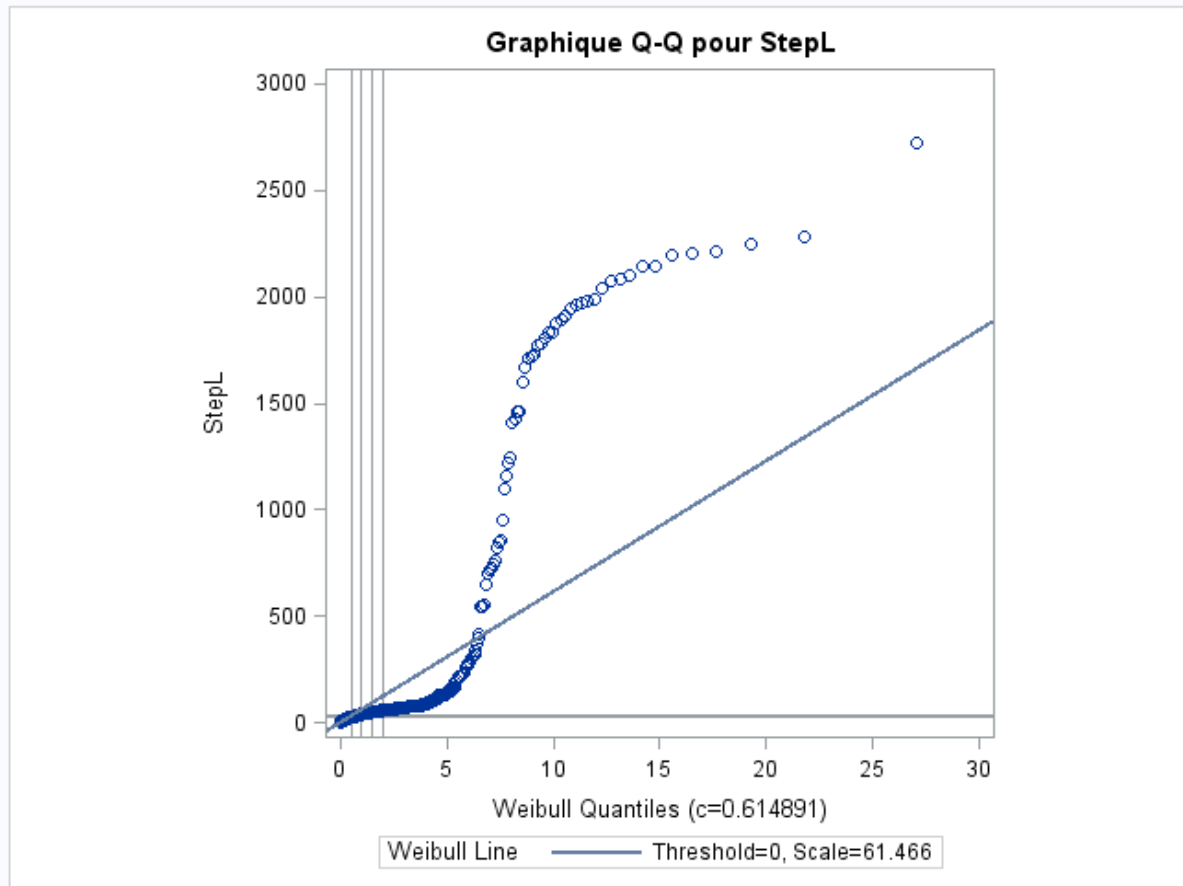


Figure 2 : diagramme quantile-quantile de la concordance des données brutes à une loi de weibull

Procédure UNIVARIATE

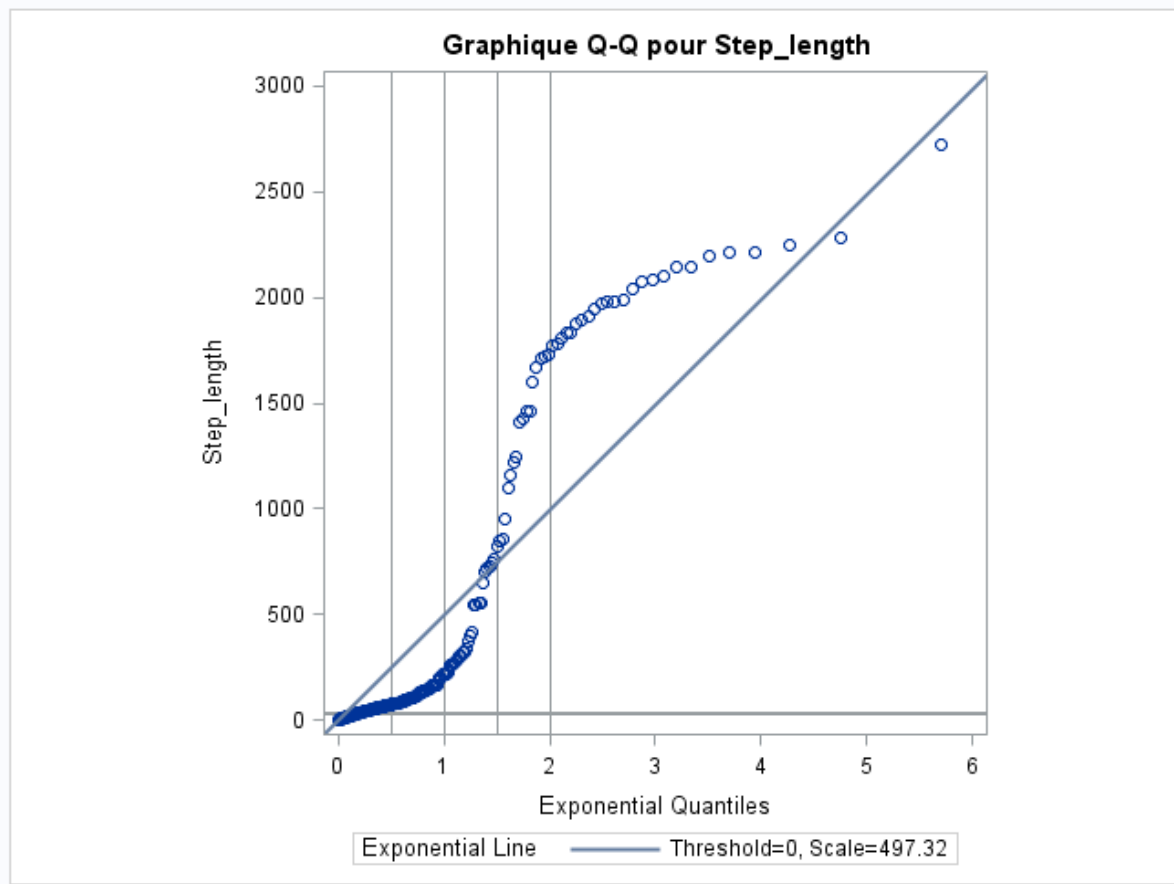


Figure 3 : diagramme quantile-quantile de la concordance des données arrondies à une loi exponentielle

Procédure UNIVARIATE

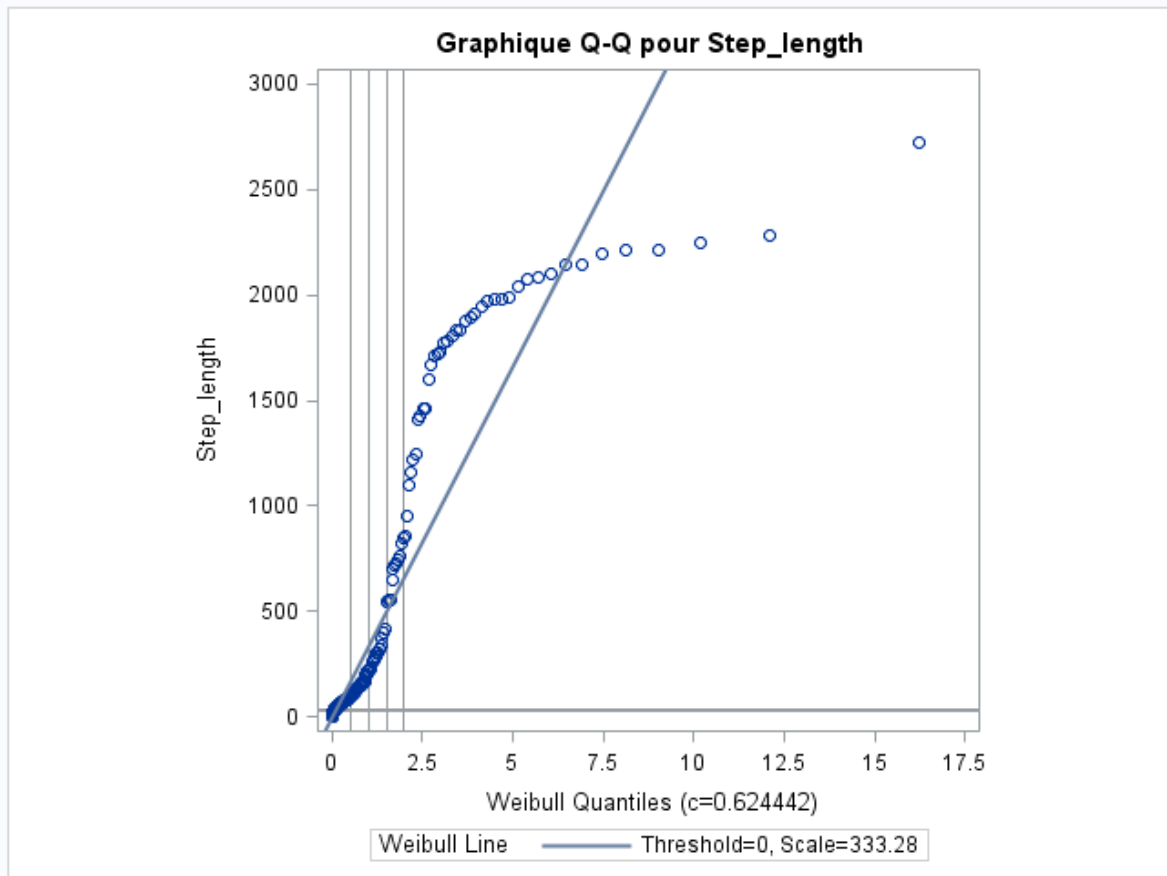


Figure 4 : diagramme quantile-quantile de la concordance des données brutes à une loi de weibull

5) DISCUSSION

Le guillemot est une espèce d'oiseaux marins et l'étude des types de mouvements de fourrages nous permet de mieux comprendre son écosystème. Le mouvement brownien correspond à un mouvement de recherche aléatoire, tandis que, le vol de Lévy correspond à une recherche optimale de l'animal. La loi de Weibull est un intermédiaire entre la loi exponentielle et la loi de type puissance : le mouvement reste aléatoire tout en ayant des pas plus longs que pour la loi exponentielle. C'est pour cela que, nous avons analysé les déplacements du guillemot entre son lieu de colonie et son lieu d'alimentation, afin d'étudier à quels types de lois semblait correspondre ces déplacements. Je me suis intéressé à principalement trois lois : loi de Weibull, loi exponentielle et loi de type puissance.

L'estimation statistique peut se faire par différentes méthodes :

- la méthode des moments,
- la méthode par maximum de vraisemblance,
- la méthode par intervalle de confiance.

Je me suis intéressé à la méthode par maximum de vraisemblance, afin d'estimer les paramètres d'une loi de Weibull et exponentielle car comme vu en cours, l'estimateur des moments pour une loi exponentielle est le même que l'estimateur de maximum de vraisemblance. Tandis que pour la loi de Weibull, elle admet des moments difficiles à calculer, je devrais, avec la méthode des moments, recourir à une estimation approchée comme pour l'estimation par maximum de vraisemblance.

Pour la loi exponentielle, l'estimateur par maximum de vraisemblance d'une telle loi est assez simple car il représente l'inverse de la moyenne empirique.

Or, pour estimer par maximum de vraisemblance les paramètres d'une loi de Weibull, il n'existe pas d'expression analytique de leurs estimations, mais on a estimé les paramètres de manière numérique.

Les résultats de ces estimations sont présentés dans les tableaux 1, 2, 3 et 4 ci-dessus. De plus, parfois l'estimation des paramètres pour la loi de Weibull, avec la fonction `fitdistr()`, ne fonctionnait pas, dans le sens où, l'estimation n'était pas faite et le message d'erreur suivant s'affichait : l'optimisation a échoué.

C'est pour cela, qu'au lieu de rentrer n'importe quelle valeur de départ dans la fonction `mle()`, nous avons décidé d'estimer dans un premier temps, les paramètres de manière graphique et cela de la manière suivante :

On a déjà vu que, la fonction de répartition d'une loi de Weibull s'écrivait :

$$F(x) = 1 - \exp\left(-\left(\frac{x}{\eta}\right)^\beta\right)$$

Pour obtenir une fonction linéaire, à partir de la fonction de répartition, on applique deux fois le logarithme à $1-F(x)$:

$$\text{Log}(-\log(|1 - F(x)|)) = \beta \log(x) + \beta \log(\eta)$$

A l'aide de la commande `lm()` du logiciel R, on détermine le coefficient directeur et l'ordonnée à l'origine de la droite des moindres carrés, on a donc $\hat{\beta}$ et $\hat{\eta}$:

$$\hat{\beta} = \text{coeffdirecteur et } \hat{\eta} = \exp\left(-\frac{\text{ordonnéealorigine}}{\hat{\beta}}\right)$$

Le test de Kolmogorov-Smirnov teste l'adéquation des données à une loi, dans le cadre du stage, on s'intéresse à deux lois, la loi exponentielle et la loi de Weibull. Les résultats de ces tests sont résumés dans les tableaux 6 et 7 ci-dessus.

On constate que pour les données brutes, le test de Kolmogorov-Smirnov est très significatif, ces données ne semblent pas suivre une loi de Weibull, ni une loi exponentielle. De plus, ceci est confirmé par les diagrammes quantile-quantile.

Pour les données arrondies, les données semblent ne pas suivre ni une loi exponentielle, ni une loi de Weibull. Là encore, ceci est confirmé par les diagrammes quantile-quantile.

En revanche, le test d'adéquation à une loi de Weibull n'est pas très significatif, c'est pour cela, que nous avons ré-échantillonné les données arrondies afin de réduire sa taille, car on sait que pour le calcul du risque de deuxième espèce, la taille de l'échantillon influe sur celui-ci.

Nous avons ré-échantillonné avec la commande `sample()` de R, en choisissant nous même la taille de l'échantillon.

La fonction `sample()` fournit un échantillon aléatoire de valeurs prises dans le vecteur donné. Nous avons choisis comme de taille de ré-échantillonnage, une taille de 100 et 500, mais pour les guillemots 6, 8, 10, 11, 13 et 14, la taille de ré-échantillonnage est de 150 car la taille de ces échantillons est inférieure à 500.

De plus, nous avons pu faire que trois simulation, car des messages d'erreur du type : l'optimisation a échoué et production de NaN (Not a Number) s'affichaient, en fait ceci nécessiterait beaucoup plus de simulations, mais les estimations et les tests réalisés ont été produits de la même manière qu'avant le ré-échantillonnage.

Voici les résultats des simulations :

Tableau 8 : première simulation dont la taille du ré-échantillonnage est égale à 100

	$\hat{\beta}$	$\hat{\eta}$
guillemot 1	0.6360611	254.6620129
guillemot 2	0.6934448	325.4048549
guillemot 3	0.7015143	416.9433511
guillemot 4	0.8789648	271.8414681
guillemot 5	0.6865055	548.7366652
guillemot 6	0.7592918	165.991515
guillemot 7	0.5823868	369.0179793
guillemot 8	0.7321393	168.5392558
guillemot 9	0.5891767	580.8713513
guillemot 10	0.5779078	0.05355844

guillemot 11	0.7469215	196.9367634
guillemot 12	0.5638738	399.4455628
guillemot 13	0.7260445	162.2307248
guillemot 14	0.7156585	125.3282151

Tableau 9 : première simulation dont la taille du ré-échantillonnage est égale à 500 sauf pour les guillemots 6, 8, 10, 11, 13 et 14 dont la taille est de 150

	$\hat{\beta}$	$\hat{\eta}$
guillemot 1	0.6260489	341.8653632
guillemot 2	0.7098608	373.0434074
guillemot 3	0.7182605	380.4047627
guillemot 4	0.872895	250.067029
guillemot 5	0.6823275	576.8930024
guillemot 6	0.6531418	189.2295692
guillemot 7	0.5813353	393.4050937
guillemot 8	0.751336	213.352573
guillemot 9	0.5770137	537.0837961
guillemot 10	0.5757335	215.7745832
guillemot 11	0.7392804	158.882727
guillemot 12	0.5662095	395.8065829
guillemot 13	0.7070629	123.285905
guillemot 14	0.6920609	133.5115941

Tableau 10 : résultats du test de Kolmogorov-Smirnov des données arrondies après ré-échantillonnage

	p-valeurs	
	loi de weibull	
guillemot 1	0.03284	0.00524
guillemot 2	0.1359	0.007255
guillemot 3	0.1722	0.004076
guillemot 4	0.7502	0.01152
guillemot 5	0.08826	0.0001654
guillemot 6	0.07669	0.002173
guillemot 7	2.24E-03	4.95E-06
guillemot 8	2.59E-02	5.41E-04
guillemot 9	2.17E-03	8.27E-06
guillemot 10	6.35E-03	2.95E-05
guillemot 11	0.03837	0.00924
guillemot 12	3.37E-03	5.86E-06
guillemot 13	0.01894	0.003324
guillemot 14	0.02603	0.002599

La première colonne représente les p-valeurs dont la taille de ré-échantillonnage est égale à 100, on peut remarquer que pour cinq des guillemots, leurs données semblent suivre une loi de Weibull, tandis que, pour le ré-échantillonnage de tailles 500 et 150 pour certains guillemots, les données semblent ne pas correspondre à une loi de Weibull.

Tableau 11 : deuxième simulation dont la taille du ré-échantillonnage est égale à 100

	$\hat{\beta}$	$\hat{\eta}$
guillemot 1	0.6431659	276.1166474
guillemot 2	0.6810885	338.2139292
guillemot 3	0.71882	384.33545
guillemot 4	0.7893447	248.899979
guillemot 5	0.6825944	537.6017446
guillemot 6	0.7680219	131.3304444
guillemot 7	0.7005516	76.0777881
guillemot 8	0.7163625	150.8177706
guillemot 9	0.4975585	134.1204598
guillemot 10	0.5928471	119.3362008
guillemot 11	0.8608016	149.0466093
guillemot 12	0.6305585	98.8081658
guillemot 13	0.7321788	120.1009245
guillemot 14	0.6987944	119.9054442

Tableau 12 : deuxième simulation dont la taille du ré-échantillonnage est égale à 500 sauf pour les guillemots 6, 8, 10, 11, 13 et 14 dont la taille est de 150

	$\hat{\beta}$	$\hat{\eta}$
guillemot 1	0.6184595	337.2705631
guillemot 2	0.6933173	325.616387
guillemot 3	0.7122658	404.1383528
guillemot 4	0.8918593	103.5136117
guillemot 5	0.6056168	166.0682444
guillemot 6	0.6848099	162.2112798
guillemot 7	0.6299888	125.5539917
guillemot 8	0.7628458	172.5386412
guillemot 9	0.5182089	215.919508
guillemot 10	0.5637117	184.7916202
guillemot 11	0.7440324	172.8246463
guillemot 12	0.5362164	245.9129256
guillemot 13	0.9503205	101.9635348
guillemot 14	0.7074677	131.5191577

Tableau 13 : résultats du test de Kolmogorov-Smirnov des données arrondies après ré-échantillonnage

p-valeurs		
loi de weibull		
guillemot 1	0.04278	0.00549
guillemot 2	0.2689	0.01331
guillemot 3	0.0972	0.001228
guillemot 4	2.95E-01	2.22E-02
guillemot 5	4.04E-02	1.39E-04
guillemot 6	0.0134	0.01196
guillemot 7	0.002476	3.68E-06
guillemot 8	0.001124	0.0007352
guillemot 9	0.002582	1.43E-07
guillemot 10	0.0002815	5.50E-05
guillemot 11	3.11E-02	9.04E-03
guillemot 12	2.82E-04	1.52E-06
guillemot 13	0.004985	0.02481
guillemot 14	0.0004751	2.73E-03

Dans cette deuxième simulation, La première colonne représente les p-valeurs dont la taille de ré-échantillonnage est égale à 100, on peut remarquer que pour seulement le deuxième guillemot, ces données semblent suivre une loi de Weibull, tandis que, pour tout le reste, les données semblent ne pas correspondre à une loi de Weibull.

Tableau 14 : troisième simulation dont la taille du ré-échantillonnage est égale à 100

	$\hat{\beta}$	$\hat{\eta}$
guillemot 1	0.6349836	324.7172103
guillemot 2	0.775978	401.536692
guillemot 3	0.7345203	477.4249536
guillemot 4	0.8381541	271.8148716
guillemot 5	0.6563848	562.6093494
guillemot 6	0.7348434	167.5778572
guillemot 7	0.6782132	87.6194342
guillemot 8	0.7401525	155.4940586
guillemot 9	0.5155441	110.0419439
guillemot 10	0.5484695	203.3038776
guillemot 11	0.9159581	157.539735
guillemot 12	0.6189912	77.7534484
guillemot 13	0.823316	118.779307
guillemot 14	0.7369133	112.9296192

Tableau 15 : troisième simulation dont la taille du ré-échantillonnage est égale à 500 sauf pour les guillemots 6, 8, 10, 11, 13 et 14 dont la taille est de 150

	$\hat{\beta}$	$\hat{\eta}$
guillemot 1	0.6263918	356.4473301
guillemot 2	0.6874286	322.1737766
guillemot 3	0.7148876	432.3994961
guillemot 4	0.7798623	129.9278049
guillemot 5	0.6043969	143.562608
guillemot 6	0.713229	171.79805
guillemot 7	0.5529054	214.5287091
guillemot 8	0.766758	158.221644
guillemot 9	0.5134007	213.8685448
guillemot 10	0.5697256	200.7328533
guillemot 11	0.7576384	172.5761766
guillemot 12	0.5310233	201.6161906
guillemot 13	0.7825112	113.1978185
guillemot 14	0.6975059	140.2568293

Tableau 16 : résultats du test de Kolmogorov-Smirnov des données arrondies après ré-échantillonnage

	p-valeurs	
	loi de weibull	
guillemot 1	0.03228	0.003744
guillemot 2	0.05431	0.01029
guillemot 3	0.2219	0.01158
guillemot 4	2.88E-01	1.22E-04
guillemot 5	2.74E-02	2.60E-05
guillemot 6	2.09E-03	1.43E-03
guillemot 7	0.01118	3.84E-07
guillemot 8	0.0004488	0.0008591
guillemot 9	0.0002858	1.13E-07
guillemot 10	6.58E-05	1.42E-05
guillemot 11	1.70E-02	7.00E-03
guillemot 12	2.77E-03	2.16E-06
guillemot 13	2.31E-02	9.55E-03
guillemot 14	0.00692	3.46E-03

Dans cette dernière simulation, La première colonne représente les p-valeurs dont la taille de ré-échantillonnage est égale à 100, on peut remarquer que pour seulement le deuxième et le troisième guillemot, leurs données semblent suivre une loi de Weibull, tandis que, pour tout le reste, les données semblent ne pas correspondre à une loi de Weibull.

J'ai remarqué que dans un premier temps, les données semblaient ne pas correspondre ni à une loi de Weibull, ni à une loi exponentielle. Puis, en ré-échantillonnant les données on a pu constater que dans certains cas, ces données semble correspondre à une loi de Weibull mais pour seulement un ré-échantillonnage de taille 100.

C'est pour cela que nous avons choisi deux critères pour sélectionner le meilleur modèle, le critère de l'AIC et du BIC.

Le critère d'information d' Akaike (AIC) est donné par la formule suivante :

$$AIC = -2\text{Log}(L) + 2k$$

où L représente la fonction de vraisemblance, k est le paramètre sur un échantillon de taille n, dans notre cas k représente le nombre de paramètre de la loi ainsi étudiée.

Nous avons calculé l'AIC, pour la loi de Weibull seulement, car mon maître de stage avait déjà calculé ce critère pour la loi exponentielle et la loi de type puissance.

Ainsi, le « meilleur » modèle sera celui dont l'AIC sera le plus petit.

Tableau 17 : résultats du critère de l'AIC

	AIC weibull	AIC puissance	AIC exponentielle
guillemot 1	5222.044	9514.82045	13317.0633
guillemot 2	5621.36	6046.70369	8880.4816
guillemot 3	7666.996	9583.75838	13670.2913
guillemot 4	4541.248	23678.3434	28356.9967
guillemot 5	5620.768	21867.244	30229.3827
guillemot 6	3378.81	2433.92502	3616.87429
guillemot 7	5770.966	10530.8707	14606.212
guillemot 8	3841.11	2644.72242	3881.48166
guillemot 9	5321.658	7375.59188	11373.2287
guillemot 10	3740.454	2803.5211	3993.67269
guillemot 11	3896.232	2861.33728	3669.28507
guillemot 12	5001.878	9105.20656	12837.6587
guillemot 13	3482.426	3437.45376	5342.47799
guillemot 14	2676.978	1694.94201	2638.44955

Pour les guillemots, 1, 2, 3, 4, 5, 7, 9 et 12, le « meilleur » modèle est celui dont les données semblent suivre une loi de Weibull car l'AIC est le plus petit. Pour les autres guillemots, on choisira le modèle qui semble suivre une loi de type puissance.

Le critère d'information de Schwartz ou BIC est donné par la formule suivante :

$$\text{BIC} = -2\text{Log}(L) + k\text{Log}(n)$$

où L représente la fonction de vraisemblance, k est le paramètre sur un échantillon de taille n, dans notre cas k représente le nombre de paramètre de la loi ainsi étudiée.

Nous avons calculé le BIC, pour la loi de Weibull, la loi exponentielle et la loi de type puissance.

Ainsi, le « meilleur » modèle sera celui dont le BIC sera le plus petit.

Tableau 18 : résultats du critère du BIC

	BIC weibull	BIC puissance	BIC exponentielle
guillemot 1	5232.31059	9525.08704	13322.1966
guillemot 2	5630.64236	6055.98606	8885.12278
guillemot 3	7677.06478	9593.82716	13675.3257
guillemot 4	4553.05815	23690.1536	28362.9018
guillemot 5	5632.53263	21879.0086	30235.265
guillemot 6	3386.18395	2441.29897	3620.56126
guillemot 7	5781.18739	10541.0921	14611.3227
guillemot 8	3848.58959	2652.20201	3885.22145
guillemot 9	5331.35189	7385.28577	11378.0757
guillemot 10	3747.88806	2810.95516	3997.38972
guillemot 11	3903.61272	2868.71799	3672.97543
guillemot 12	5011.88231	9115.21087	12842.6608
guillemot 13	3490.74842	3445.77618	5346.6392
guillemot 14	2683.8279	1701.79191	2641.8745

Les résultats sont les même que pour le critère de l'AIC.

Pour les guillemots, 1, 2, 3, 4, 5, 7, 9 et 12, le « meilleur » modèle est celui dont les données semblent suivre une loi de Weibull car le BIC est le plus petit. Pour les autres guillemots, on choisira le modèle qui semble suivre une loi de type puissance.

Bien que le mouvement brownien corresponde à une recherche aléatoire, on ne peut conclure que le guillemot adopte ce mouvement lors de sa recherche alimentaire.

De plus, le vol de Lévy qui représente une recherche alimentaire optimale de l'animal, on ne peut là aussi conclure que le guillemot adopte ce mouvement lors de sa recherche alimentaire.

Néanmoins, les différents critères de sélection, nous conduisent à déduire que pour certains guillemots le « meilleur » modèle est celui dont les données semblent suivre une loi de type puissance, on en déduit que pour ces guillemots, le vol de Lévy semble être la stratégie de leur recherche alimentaire.

Pour la suite, il reste à déterminer si l'âge et le poids ont une influence sur les déplacements car les différents guillemots étudiés ont des âges et des poids différents. Etant donné que la collecte des données des déplacements a été réalisée toute les deux minutes, il faudrait utiliser l'ANOVA pour plans à mesures répétées. En effet, plus l'animal est âgé et plus son déplacement vers les lieux d'alimentation est dirigé ce qui correspondrait à l'utilisation du vol de Lévy. De plus, en adoptant cette stratégie de recherche alimentaire, l'animal augmente son poids, car il mange plus de proies en volant moins, qu'un individu qui fait du vol brownien.

6) Bibliographie

POULIN, N. (2011-2012). Master 1 Cours d'Estimation statistique

DORTET, J.-L. (2011-2012). Master 1 Cours de Test statistique

BERTRAND, F. (2011-2012). Master 1 Cours de Plans d'expériences, *Quelques représentations graphiques usuelles des échantillons*

BERTRAND, F. (2011-2012). Master 1 Cours de Plans d'expériences, *Tests de normalité Distributions univariées et multivariées*

GOVAERTS, B. Cours Estimation par maximum de vraisemblance Approche numérique.

RICCI, V. (2005) : <http://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf>

MORICE, E. (1968). Quelques problèmes d'estimation relatifs à la loi de Weibull. (*Revue de Statistique Appliquée*)

SUEUR, C., BRIARD, L., PETIT, O. (2011). Individual Analyses of Lévy Walk in Semi-Free Ranging Tonkean Macaques (*Macaca tonkeana*), *PLoS ONE* 6(10): e26788. doi :10.1371/journal.pone.0026788

BARTUMEUS et al. , Fishery Discards Impact on Seabird Movement Patterns at Regional Scales, *Current Biology* (2010), doi:10.1016/j.cub.2009.11.073

EDWARDS, A.M., (2008). Using likelihood to test for Lévy flight search patterns and for general power-law distributions in nature, *Journal of Animal Ecology* doi : 10.1111/j.1365-2656.2008.01428.x

MARELL, A., BALL J.P., HOFGAARD A. (2002). Foraging and movements paths of female reindeer : insights from fractal analysis, correlated random walks, and Lévy flights, *NRC Research Press* DOI : 10.1139/Z02-061

MIRAMONTES, O., BOYER, D., BARTUMEUS, F. (2012). The Effects of Spatially Heterogeneous Prey Distributions on Detection Patterns in Foraging Seabirds, *PLoS ONE* 7(4): e34317. doi:10.1371/journal.pone.0034317.

REYNOLDS, A.M., (2012). Olfactory search behaviour in the wandering albatross is predicted to give rise to Lévy flight movement patterns. *Animal Behaviour* doi:10.1016/j.anbehav.2012.02.014.

SUEUR, C. (2011). A Non-Lévy Random Walk in Chacma Baboons : What Does It Mean ? *PLoS ONE* 6(1): e16131. doi :10.1371/journal.pone.0016131