

---

---

# Table des matières

---

<b>1 ASPECTS TECHNIQUES SUR LES MOUSTIQUAIRES ET INSECTICIDES EVALUES</b>	<b>v</b>
1.1 SITE D'EXPERIMENTATION DES MOUSTIQUAIRES IMPREGNEES . . .	v
1.2 INSECTICIDES EVALUES . . . . .	v
1.3 MOUSTIQUAIRES UTILISEES . . . . .	vi
1.4 TECHNIQUES D'IMPREGNATION DES MOUSTIQUAIRES . . . . .	vi
1.4.1 ECHANTILLONNAGE . . . . .	vii
1.4.2 TEST D'IRRITABILITE . . . . .	vii
<b>2 LES METHODES</b>	<b>viii</b>
2.1 ESTIMATION DES PARAMETRES D'UNE LOI . . . . .	viii
2.1.1 LA METHODE DU MAXIMUM DE VRAISEMBLANCE . . . . .	viii
2.1.2 EXEMPLES . . . . .	viii
2.1.3 LA METHODE DES MOMENTS . . . . .	x
2.1.4 METHODE D'ESTIMATION BASEE SUR LA FONCTION DE REPARTITION EMPIRIQUE . . . . .	x
2.2 L'ANALYSE DE SURVIE . . . . .	xi
2.2.1 DEFINITIONS ET EXEMPLES . . . . .	xi
2.2.2 OUTILS PROBABILISTES DE L'ANALYSE DE SURVIE . . . . .	xi
2.3 ESTIMATION NON PARAMETRIQUE . . . . .	xiv
2.4 LE TEST DU LOG-RANK . . . . .	xv
2.4.1 PROCESSUS DE COMPTAGE . . . . .	xv
2.4.2 LES MARTINGALES . . . . .	xvi
<b>3 APPLICATIONS DES METHODES ET RESULTATS</b>	<b>xx</b>
3.1 INTRODUCTION . . . . .	xx
3.2 RECHERCHE DE LA LOI . . . . .	xx
3.3 ESTIMATION DES AUTRES PARAMETRES . . . . .	xxii
3.4 PRESENTATION DES RESULTATS DE L'ANALYSE DE SURVIE . . . . .	xxii

---

---

# DEDICACES

---

A mes parents,  
Monsieur et Madame **BESSALA**

Papa, Mamman,  
je vous remercie pour tout l'amour que vous m'avez donné.  
je vous remercie pour tous vos sacrifices.  
je vous remercie pour votre soutien, tant sur le plan moral que financier.  
je vous remercie pour l'éducation que vous m'avez donnée.  
voyez en ce mémoire, le premier fruit de vos efforts consentis à l'endroit de votre fils que je suis.

A mes soeurs,

NGA Laurence Rosine, MENYE Calixte Nadine, NGONO Marie Louise, BESSALA Prudence  
Valérie, AYISSI Ferdinande Christelle, NGONO Marie Nadine, BILOA Dorothée Mireille.

Merci pour votre indéniab le soutien, pour votre disponibilité et surtout pour l'amour que  
vous manifestez envers moi quotidiennement.

A mon grand-frère et ami GuY Severin MAHIANE,

Merci pour toute cette peine prise depuis près de dix ans aujourd'hui, sans toi, ma passion  
pour les mathématiques n'aurait pas vu le jour; ceci est un fruit de tous ces efforts.

A DIOP Salimata,

Tu es un être qui occupe une place de choix dans ma vie, je te remercie infiniment pour  
tous les sacrifices faits pour moi, pour l'amour vrai que tu m'as toujours porté; vois en ce  
mémoire le début d'une vie future. Tu es et resteras précieuse au plus profond de moi.

---

# REMERCIEMENTS

---

Au Pr Henri GWET

Pour le temps et l'investigation consacrés tout au long de mon stage.

Au Dr Josiane ETANG

Pour sa disponibilité et ses conseils malgré ses occupations

Au Dr Eugène-Patrice NDONG NGUEMA

Pour les conseils judicieux qui ont fortement contribué à la résolution du problème posé lors du stage.

Au Cabinet Administratif de l'OCEAC

Pour m'avoir permis d'effectuer ce stage dans le cadre de la réalisation de mon mémoire.

Au Pr Didier Dacunha-Castelle

Pour avoir initié le Master de Statistiques au Cameroun

A tous les professeurs du Master de Statistiques

Les enseignements que vous nous avez prodigués ont un contenu riche et nous permettent d'être des statisticiens de haut niveau.

A tous mes camarades du Master de Statistiques

Pour avoir toujours été disponibles quand j'ai fait appel à eux.

A toute la promotion 2001 de la Terminele C du Lycée de Mballa II

Vous avez tous été des camarades solidaires, vous me manquez tous.

A mon oncle: AYISSI AYISSI Patrice

Pour le soutien que tu ne cesses de m'apporter; tu restes un homme idéal pour moi.

A mes amis: Guy Severin MAHIANE, Richard BAFAKAN, HOUSSEINI Bouba, NGOME

BISSENE Yannick, SOP Tresor, ETUNE Jean-Jacques, Mamadou NDOYE, Sims

LUTUMBA, Claude BANA, Patrick EKANI, MBOMBOA Fatima, Tata ADAMA, Saoud

ABDESLAM, Lahcen NOUARRI, Naima ZIOUH, Joel LOKOTA, SAIBOU

Abdoulahi, TCHASSEP Junior, CHEGUE Isaac, BASSANGA Modeste, Pierre

ELOUNDOU, NYAMEN Linda et KUEMO ARMELLE.

Merci pour tout ce que vous m'avez apporté.

A Mlle Pauline ZOBO

Merci pour l'équilibre que tu m'apportes, pour ta sympathie et ta disponibilité.

---

---

# INTRODUCTION

---

La fin de la seconde guerre mondiale est marquée par la découverte d'insecticides qui ont pesé de façon importante dans la lutte chimique contre les insectes nuisibles en agriculture et en santé publique, notamment dans la lutte contre le paludisme. La malaria a ainsi pu être éradiquée dans plusieurs parties du globe, grâce aux aspersions intradomiciliaires. En Afrique tropicale, après avoir pendant quelques années donné de réels espoirs de voir le paludisme éliminé du continent, les aspersions intradomiciliaires ont dûes être interrompues dans la zone équatoriale pour cause entre autre d'apparitions de résistance des anophèles aux insecticides. Depuis une vingtaine d'années, les moustiquaires imprégnées d'insecticides sont expérimentées et maintenant recommandées dans la lutte contre le paludisme. Les insecticides utilisés font cependant malheureusement face à quelques difficultés: développement de la résistance aux insecticides, effets secondaires sur les utilisateurs, faible persistance de l'efficacité après lavages de certaines formulations. Face à ces problèmes, l'OMS encourage la recherche et l'évaluation de nouveaux insecticides en tant que principes actifs ou de nouvelles formulations, pour disposer d'une gamme élargie de produits commerciaux pouvant répondre de façon durable aux besoins de la lutte contre les insectes, vecteurs de maladies. Dans ce sillage, plusieurs méthodes ont été développées dans le but de comparer de pouvoir comparer les efficacités relatives des insecticides; ainsi, JM Hougard et al(2003) ont proposé une méthode reposant sur le caractère temporel de la variable mesurée, en introduisant des techniques de l'analyse de survie(représentation de Kaplan-Meier) plus adaptées pour des données censurées. La mesure de l'efficacité d'un insecticides est fonction de plusieurs paramètres, celui sur lequel le cadre de notre étude porte est le critère d'irritabilité. Ce faisant, d'un point de vue conceptuel, la démarche méthodologique que nous envisageons ici repose sur la méthode proposée par JM Hougard et al. Ceci étant, nous répondrons au problème posé après avoir d'une part présenté les aspects théoriques des différentes méthodes utilisées, en contournant des problèmes ponctuels intervenus au cours du stage.

# ASPECTS TECHNIQUES SUR LES MOUSTIQUAIRES ET INSECTICIDES EVALUES

---

## 1.1 SITE D'EXPERIMENTATION DES MOUSTIQUAIRES IMPREGNEES

Les moustiquaires imprégnées dont l'efficacité a été évaluée au cours de l'étude ont été soumises aux conditions naturelles d'utilisation dans le village Ebogo pendant la période de Mai à Novembre 1998.

Le village Ebogo est bâti à 15 kilomètres environ au sud de Mbalmayo, entre la rive gauche du Nyong et la rive droite du So'o. Les deux cours d'eau constituent des gîtes permanents à *An.moucheti*, le principal vecteur du paludisme dans ce site. La transmission du paludisme y est permanente, environ 307 piqûres infectantes par homme et par an, avec un pic en saison des pluies; *An.gambiae* y transmet également le paludisme, environ 48 piqûres infectantes par homme et par an (Njan Nloga et al., 1993). Ce site a été choisi pour l'étude, parce qu'il avait déjà abrité un programme d'évaluation de l'efficacité des moustiquaires en 1992 (Njan Nloga et al., 1993). Les habitants de ce village étaient habitués à dormir sous moustiquaires imprégnées, ce qui a été un acquis pour les besoins de sensibilisation et de faisabilité de l'étude.

## 1.2 INSECTICIDES EVALUES

Les insecticides dont l'irritabilité a été évaluée sont: la perméthrine, la deltaméthrine et la cyfluthrine.

### a-La perméthrine

Elle est le premier pyrethrineoïde à être utilisé en santé publique; elle constitue à cet effet un insecticide de référence. La formulation employée dans l'étude (Emulsifiable Concentrate ou EC) est une émulsifiable laiteuse; la dose d'imprégnation a été de  $500 \text{ mgm.a/m}^2$

### **b-La deltaméthrine**

C'est un alpha-cyanopyrethriinoïde couramment utilisé au Cameroun et adopté par le programme national de lutte contre le paludisme pour l'imprégnation des moustiquaires. Comme tous les alpha-cyanopyrethriinoïdes, elle agit à une dose beaucoup plus faible que la perméthrine, soit 15 à 25  $mg.m.a/m^2$ . Les formulations utilisées dans l'étude sont:

-La formulation "Suspension Concentrée ou SC" qui est suspension stable de matière active, la dose d'imprégnation étant de 20  $mg.m.a/m$

-La formulation en tablettes "K'OTAB", qui est une combinaison de l'insecticide avec une poudre jaunâtre, sous forme de comprimés solubles dans l'eau (400 mg de matière active par comprimé); la dose d'imprégnation a été de 25  $mg.m.a/m^2$

### **c-La cyfluthrine**

En dehors des insecticides cités précédemment, un autre pyréthriinoïde souvent utilisé pour l'imprégnation est la cyfluthrine.

## **1.3 MOUSTIQUAIRES UTILISEES**

Les moustiquaires utilisées ont été importées de Thaïlande; elles ont été fabriquées par la société SIAMDUCTH; elles étaient de couleur blanche, en fibres de polyester et en forme de parallélepède-rectangle, de texture "75 denier, 156 mesh (12\*36)". Le denier représente la consistance de la fibre; la mesh ou maille correspond au nombre de boucles que forme la fibre sur  $625cm^2$  de tulle moustiquaire. Les grandes moustiquaires (X-family) mesuraient  $16.30 m^2$  de surface et petites (family) mesuraient  $13.13m^2$

## **1.4 TECHNIQUES D'IMPREGNATION DES MOUSTIQUAIRES**

### **a-Test de rétention d'eau**

Ce test permet de calculer le volume de la suspension insecticide-eau nécessaire pour l'imprégnation d'un type donné de moustiquaires. Il consiste à immerger une moustiquaire de surface donnée dans un volume d'eau connu  $V_1$ , à l'essorer et à mesurer le volume d'eau restant  $V_2$ ; le volume d'eau retenu par la moustiquaire est alors déterminé par la relation:

$$V_x = V_1 - V_2$$

Pour une bonne estimation, il est nécessaire de faire trois essais, puis de calculer le volume moyen  $V_m$  de rétention. Le volume d'eau déterminé est alors divisé par la surface de la moustiquaire, pour évaluer la rétention d'eau en  $ml/m^2$  de tulle moustiquaire.

## **b-Calcul des doses et trempage**

Les moustiquaires ont été imprégnées par la méthode de trempage("dipping"). La quantité d'insecticide requise(mg) pour l'imprégnation a été calculée en multipliant la dose cible (en matière active/ $m^2$ ). Cette quantité (en mg) a été convertie en volume équivalent(ml), puis le volume équivalent a été complété avec de l'eau pour atteindre le volume total estimé par le test de rétention d'eau. Ainsi, les solutions d'imprégnation ont été préparées pour des lots de 10 moustiquaires, en multipliant par 10 les volumes d'eau et d'insecticide prévus pour une moustiquaire.

Les moustiquaires préalablement numérotées ont été trempées en lots de 10, essorées et séchées sur une surface plane.

Des bio-essais ont été réalisés sur des moustiquaires nouvellement imprégnées et non utilisées, puis sur des moustiquaires en cours d'utilisation prélevées sur le terrain. Les anophèles utilisées étaient des ascendants d'une souche d'*An.gambiae* élevée au laboratoire: la souche sensible de référence(souche Kisumu)

### **1.4.1 ECHANTILLONNAGE**

Juste après l'imprégnation, 8 moustiquaires imprégnées ont été choisies au hasard (2 par insecticide) pour évaluer l'irritabilité immédiate. Par la suite, 8 autres moustiquaires ont été retirées des maisons chaque mois pendant 6 mois, pour tester la rémanence de l'insecticide. Au laboratoire, les moustiquaires ramenées du terrain ont été identifiées par leurs numéros d'ordre, puis des carrés(25\*25cm) y ont été prélevées sur le bord inférieur. Tous les échantillons ont été étiquetés selon le numéro de la moustiquaire, sa durée d'utilisation, l'insecticide en imprégnation, puis emballés dans du papier aluminium et conservés au réfrigérateur(4°C) jusqu'au jour du test.

### **1.4.2 TEST D'IRRITABILITE**

Les moustiques sont emprisonnés à l'intérieur d'un cône dont la partie inférieure est couverte par un tulle moustiquaire imprégné d'insecticide; les moustiques bien qu'emprisonnés dans le cône ont la possibilité de se poser sur les parois de celui-ci; ces parois sont glissantes, de sorte que le moustique qui est emprisonné soit toujours en contact avec le tulle s'il n'est pas en vol; l'effet irritant de l'insecticide peut amener les moustiques à éviter le contact; le test est réalisé en chronométrant dans un intervalle de trois minutes, le temps réel que le moustique passe au contact du tulle moustiquaire. Un seul moustique est introduit dans le cône; l'on note alors le temps du premier envol de ce moustique après soixante secondes; si le moustique passe plus de 256 secondes sans s'envoler, on conclut que l'insecticide n'a pas un effet irritant. Pour chaque insecticide, cette opération est reprise 50 fois avec 50 moustiques différents sur deux moustiquaires imprégnées de cet insecticide.

# LES METHODES

---

L'objectif dans ce chapitre est de présenter l'aspect théorique des différentes techniques statistiques que nous allons appliquer aux méthodes par la suite; c'est ainsi que nous parlerons d'abord des différentes méthodes d'estimation statistique, nous abordons ensuite les techniques de l'analyse de survie; nous allons nous intéresser au cas non paramétrique.

## 2.1 ESTIMATION DES PARAMETRES D'UNE LOI

### 2.1.1 LA METHODE DU MAXIMUM DE VRAISEMBLANCE

Soit  $p \geq 1$ . On suppose le modèle  $(x, A, P_\theta)$   $\theta$  appartenant à  $\Theta \subseteq \mathbb{R}^p$  dominé par une mesure  $\mu$   $\sigma$ -finie. On note  $f(x, \theta)$  la densité de  $P_\theta$  par rapport à  $\mu$  et on suppose que  $\Theta$  est un ouvert de  $\mathbb{R}^p$ .

**Définition 2.1.** On définit la fonction de vraisemblance ou vraisemblance au point  $x$  par:

$$L_x: \Theta \longrightarrow \mathbb{R}_+$$

$$\theta \longmapsto L_x(\theta) = f(x, \theta).$$

**Définition 2.2.**  $\hat{\theta}(X)$  est un estimateur du maximum de vraisemblance (EMV) de  $\theta$  si:

$$f(x, \hat{\theta}(x)) = \sup_{\theta \in \Theta} f(x, \theta) \quad \mu\text{-pp.}$$

**Remarque 2.1.** Si  $f(x, \theta) \geq 0$   $\mu$ -pp  $\forall \theta$  et si  $f(x, \theta)$  est 2 fois continûment différentiable en  $\theta$ ,  $\hat{\theta}(X)$  est un EMV de  $\theta$  ssi:

$$\text{grad} \ln(f(x, \theta)) = 0 \text{ et } \left( \frac{\partial^2 \ln(f(x, \theta))}{\partial \theta_i \partial \theta_j} \right) \Big|_{\hat{\theta}(x)} \text{ est semi-négative.}$$

L'ensemble constitué par ces deux équations est appelé "équations de vraisemblance".

### 2.1.2 EXEMPLES

#### LOI NORMALE

On dit que  $X$  suit une loi normale de moyenne  $m$  et de variance  $\sigma^2$  et on note  $X \hookrightarrow N(m, \sigma^2)$  si:

$$f_X(x, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x_i - m}{\sigma}\right)^2\right) \text{ où } \theta = (m, \sigma^2). \text{ Déterminons l'EMV de } \theta.$$

Supposons observées  $n$  réalisations de la loi de  $X$ ; on dispose alors d'un échantillon de taille

n de la loi de X.

On a:

$$f_{X_i}(x_i, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x_i - m}{\sigma}\right)^2\right)$$

ainsi pour x dans  $\mathbb{R}^n$ , on a la vraisemblance:

$$L_x(\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x_i - m}{\sigma}\right)^2\right). \text{ Il vient:}$$

$$\ln(L_x(\theta)) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - m}{\sigma}\right)^2.$$

$$\text{Dès lors } \text{grad} \ln(L_x(\theta)) = 0 \iff: \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - m) = 0 \text{ et } \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - m)^2 = 0$$

$$\iff \hat{m} = \frac{1}{n} \sum_{i=1}^n (X_i) \text{ et}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{m})^2$$

c'est à dire:

$$\hat{\theta}(X) = (\hat{m}, \hat{\sigma}^2)$$

On vérifie facilement que la matrice:

$$\left( \frac{\partial^2 \ln(L_x(\theta))}{\partial \theta_i \partial \theta_j} \right) |_{\hat{\theta}(X)} \text{ est semi-négative}$$

## LOI LOG-NORMALE

$X \hookrightarrow \text{log-N}(m, \sigma^2)$  si:

$Y = \ln(X)$  suit une loi normale de paramètres  $(m, \sigma^2)$

Par le maximum de vraisemblance en procédant comme précédemment, on obtient les estimateurs suivants:

$$\hat{m} = \frac{1}{n} \sum_{i=1}^n (\ln(X_i)) \text{ et}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (\ln(X_i) - \hat{m})^2$$

**Remarque 2.2.** *Il n'est pas toujours possible d'obtenir une forme explicite pour l'EMV de  $\theta$*

## EXEMPLE

On dit qu'une variable aléatoire X suit une loi de Weibull standard si sa fonction de répartition est:

$$F_X(x) = \exp(-x^\theta) \text{ avec } \theta > 0$$

Sa densité est alors:

$$f_X(x, \theta) = \theta x^{\theta-1} \exp(-x^\theta)$$

L'estimation du paramètre  $\theta$  par la méthode du maximum de vraisemblance nous donne:

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n (x_i^{\hat{\theta}} - 1) \ln(x_i)}.$$

Cette équation ne peut se résoudre que numériquement par des approximations successives

### 2.1.3 LA METHODE DES MOMENTS

Elle consiste à prendre comme paramètre de l'estimateur de  $\theta$  la solution dont on suppose l'existence et l'unicité du système suivant:

$U_i(n) = m_i(\theta)$   $i = 1, 2, \dots, p$  avec les  $U_i$  sont les moments empiriques d'ordre  $i$  et les  $m_i$  les moments théoriques d'ordre  $i$

#### EXEMPLE

Une variable aléatoire  $X$  suit une loi gamma de paramètres  $(a, \lambda)$  et on note  $X \hookrightarrow Ga(a, \lambda)$  si

$$f_X(x, a, \lambda) = \frac{x^{a-1} \exp\left(\frac{-x}{\lambda}\right)}{\lambda^a \Gamma(a)} \text{ si } x > 0 \text{ et}$$

$$f_X(x, a, \lambda) = 0 \text{ si non}$$

$$\text{avec } \Gamma(p) = \int_0^\infty x^{p-1} \exp(-x) dx \quad \forall p > 0$$

Ainsi donc, l'estimateur par la méthode des moments de  $(a, \lambda)$  s'obtient en appliquant le principe décrit plus haut. On obtient:

$$\hat{a} = \frac{\overline{X^2}}{\overline{X^2} - \overline{X}^2} \text{ et } \hat{\lambda} = \frac{\overline{X^2} - \overline{X}^2}{\overline{X}} \text{ où } \overline{X^2} = \frac{1}{n} \sum_{i=1}^n X_i^2 \text{ et } \overline{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

### 2.1.4 METHODE D'ESTIMATION BASEE SUR LA FONCTION DE REPARTITION EMPIRIQUE

Il arrive souvent qu'après certaines enquêtes, l'on perde les données recueillies, mais que l'on dispose de données regroupées par intervalles.

En supposant que les données "brutes" sont issues d'une loi de probabilité  $X$  de paramètre  $\theta$ , les données regroupées peuvent nous fournir la fonction de répartition empirique qu'on notera ici  $H_n$  qui estime la fonction de répartition théorique  $F_\theta$ ; on peut dès lors penser que simuler les données suivant la loi  $X$  peut suffire pour réaliser l'étude sans appréhension, mais alors une question digne d'intérêt surgit:

Quelle peut être la "meilleure" fonction de répartition théorique? en d'autres termes, quelle est le paramètre  $\theta$  qui rend  $F_\theta$  le plus proche possible de  $H_n$ ? ou encore quel est le paramètre  $\theta$  qui minimise l'écart entre  $F_\theta$  et  $H_n$

On est alors en plein dans un problème d'optimisation; ce qui peut se formuler ainsi:

Trouver  $\theta$  qui minimise  $d(F_\theta, H_n)$

Dans notre cas, nous avons choisi la distance euclidienne:

$$d(F_\theta, H_n) = \sqrt{\sum_{i=1}^n (F_\theta(x_i) - H_n(x_i))^2}$$

Il suffit donc d'implémenter le code qui minimise cette distance.

## 2.2 L'ANALYSE DE SURVIE

### 2.2.1 DEFINITIONS ET EXEMPLES

#### VARIABLE DE SURVIE

Une variable de survie est à peu près comme une variable quantitative, mais qui peut prendre des valeurs numériques précises comme "13ans", mais aussi des valeurs moins précises du genre "plus de 15 ans".

#### CENSURE

Imaginons que l'on veuille observer chez des patients un évènement sur une période donnée. Soit  $\tau$  le temps durant lequel on souhaite observer tous les patients.

On qualifie donc de "valeur censurée" des valeurs qui ne fournissent pas des informations précises.

#### EXEMPLE

Pour un individu dont la guérison n'a pas été observée en  $\tau$ , on dira qu'on a une censure "fixe".

Pour un individu qu'on a perdu de vue, on dira qu'on a une censure "aléatoire".

### 2.2.2 OUTILS PROBABILISTES DE L'ANALYSE DE SURVIE

Considérons la variable aléatoire de survie  $T$  supposée continue qui prendra ses valeurs dans  $[0, +\infty[$ .

#### SES CARACTERISTIQUES

Sa fonction de répartition

$$F_T(t) = P(T \leq t)$$

Sa densité

$$f_T(t) = F_T'(t)$$

Sa fonction de survie

$$S_T(t) = 1 - F_T(t)$$

Par exemple, si  $T$  désigne le nombre d'années de survie après le diagnostic d'un cancer,  $S_T(t)$  est la probabilité de survivre au moins  $t$  années.

On a aussi que  $S_T(t) = P(T > t)$

Sa fonction de risque instantané

Elle est définie par :

$$\lambda_T(t) = \lim_{dt \rightarrow 0} \frac{P(t < T \leq t + dt / T > t)}{dt}$$

C'est la probabilité instantanée d'un évènement juste après  $t$  sachant que cet évènement n'était pas arrivé avant  $t$ .

On a:

$$\lambda_T(t) = \lim_{dt \rightarrow 0} \frac{P(t < T \leq t + dt / T > t)}{dt} \text{ lorsque } dt \rightarrow 0$$

$$= \lim_{dt \rightarrow 0} \frac{P(t < T \leq T + dt, T > t)}{dt P(T > t)} \text{ lorsque } dt \rightarrow 0$$

$$= \lim_{dt \rightarrow 0} \frac{P(t < T \leq t + dt)}{dt S_T(t)} \text{ lorsque } dt \rightarrow 0$$

d'où

$$\lambda_T(t) = \frac{f_T(t)}{S_T(t)}$$

$$\text{Or } f_T(t) = F_T'(t)$$

$$\implies \lambda_T(t) = \frac{F_T'(t)}{S_T(t)}$$

$$\implies \lambda_T(t) = -\frac{S_T'(t)}{S_T(t)}$$

$$\implies \lambda_T(t) = -(\ln(S_T))' \text{ d'où}$$

$$S_T(t) = \exp\left(-\int_0^t \lambda_T(u) du\right)$$

Cette relation nous donne le lien entre  $S_T$  et  $\lambda_T$

Sa fonction de risque cumulé

Elle est définie par:

$$\Lambda_T(t) = \int_0^t \lambda_T(u) du$$

## EXEMPLES

### LOI EXPONENTIELLE

Densité:  $f_T(t) = \lambda \exp(-\lambda x)$  si  $x > 0$  et

$f_T(t) = 0$  si non

Fonction de survie:  $F_T(t) = \exp(-\lambda t)$

Fonction de risque instantané:  $\lambda_T(t) = \lambda = cte$

### LOI DE WEIBULL

Fonction de survie  $S_T(t) = \exp(-\lambda t^\alpha)$  avec  $\alpha > 0, \lambda > 0, t > 0$

Densité:  $f_T(t) = \lambda \alpha t^{\alpha-1} \exp(-\lambda t^\alpha)$

## FORMALISME

Soit  $\tau$  la durée au cours de laquelle on observe les sujets

Pour un individu (ième) non censuré, on a observé  $T_i = \min(T_i, \tau)$

Si l'individu est censuré fixe, on a observé  $\min(\tau, T_i)$

Si l'individu est censuré aléatoire (décès, déménagement...), on observe  $\min(T_i, C_i) = C_i$ .  
 Donc, d'une manière générale, soit  $X_i = \min(T_i, C_i)$ . Lorsque qu'on a la censure, on observe  $\min(T_i, C_i) = C_i$ ; lorsqu'on n'a pas de censure on observe  $\min(C_i, T_i) = T_i$  avec la convention suivante:

si  $\tau \leq C_i$  on a observé  $C_i$

La nature censure ou évènement d'intérêt est alors connue, on a le code:

1 si  $T_i \leq C_i$  et 0 si  $T_i > C_i$

On représentera donc une durée par un couple  $(X_i, \Delta_i)$  où :

$X_i = \min(T_i, C_i)$

et  $\Delta_i = 1_{T_i \leq C_i}$

## LOIS DE PROBABILITE

$$P(X \leq t, \Delta = 1) = P(T \leq t, T \leq C)$$

$$= \int_{u=0}^t \int_{v=u}^{\infty} f_{T,C}(u,v) du dv$$

on fait l'hypothese de censure independante, ie  $C \perp T$  d'où

$$P(X \leq t, \Delta = 1) = \int_{u=0}^t f_T(u) S_C(u) du$$

$$\text{car } S_C(u) = \int_{v=u}^{\infty} f_C(v) dv$$

d'où

$$P(X \leq t, \Delta = 1) = \int_0^t f_T(u) S_C(u) du$$

de même, on montre que:

$$P(X \leq t, \Delta = 0) = \int_0^t S_T(u) f_C(u) du$$

En dérivant par rapport à t, on obtient:

$$(X = T, 1) \longrightarrow f_T(x) S_C(x) \text{ et}$$

$$(X = C, 0) \longrightarrow S_T(x) f_C(x) \text{ d'où la vraisemblance}$$

$$L(\lambda) = \left[ \prod_{i=1}^n f_T(X_i) S_C(X_i) \right]_{\Delta_i=1} \left[ \prod_{i=1}^n S_T(X_i) f_C(X_i) \right]_{\Delta_i=0}$$

$$= \prod_{i=1}^n [f_T(X_i) S_C(X_i)]^{\Delta_i} [S_T(X_i) f_C(X_i)]^{1-\Delta_i}$$

si on suppose la censure non informative, ie la loi de C ne dépend pas des paramètres de la loi de T, on considèrera alors notre vraisemblance comme suit:

$$L(\lambda) = \prod_{i=1}^n f_T(X_i)^{\Delta_i} S_T(X_i)^{1-\Delta_i}$$

## EXEMPLE

Supposons qu'on a observé n lampes dont la durée de vie suit une loi exponentielle de paramètre  $\lambda > 0$

Pour un échantillon

$((X_1, \Delta_1), (X_2, \Delta_2), \dots, (X_n, \Delta_n))$  on a:

$$L(\lambda) = \prod_{i=1}^n [\lambda \exp(-\lambda X_i)]^{\Delta_i} [\exp(-\lambda X_i)]^{1-\Delta_i}$$

$$= \prod_{i=1}^n \lambda^{\Delta_i} \exp(-\lambda X_i)$$

$$= \lambda^{\sum_{i=1}^n \Delta_i} \exp\left(-\sum_{i=1}^n X_i\right)$$

d'où la log-vraisemblance  $l(\lambda) = \ln(L(\lambda)) = \ln(\lambda) \sum_{i=1}^n \Delta_i - \lambda \sum_{i=1}^n X_i$

ainsi,

$$l'(\lambda) = 0 \iff$$

$$\widehat{\lambda}_n^C = \frac{\sum_{i=1}^n \Delta_i}{\sum_{i=1}^n X_i}$$

$\widehat{\lambda}_n^C$  est l'estimateur censuré de  $\lambda$ ; par maximum de vraisemblance, on trouve qu'en absence de censure:

$$\widehat{\lambda}_n = \frac{n}{\sum_{i=1}^n X_i}$$

**Théorème 2.1.**  $\widehat{\lambda}_n$  est meilleur que  $\widehat{\lambda}_n^C$

**Remarque 2.3.** Supposons que  $T \hookrightarrow \varepsilon(\lambda)$  et que  $C \hookrightarrow \varepsilon(\theta)$ ; alors:

$$X = \min(T, C) \hookrightarrow \varepsilon(\lambda + \theta)$$

En effet,

$$P(X > t) = P(T > t, C > t) = P(T > t)P(C > t) \text{ car } T \perp C$$

$$= S_C(t)S_T(t)$$

$$= \exp(-(\lambda + \theta)) \text{ d'où le résultat}$$

## 2.3 ESTIMATION NON PARAMETRIQUE

### MOTIVATION

On dispose de notre échantillon  $((X_1, \Delta_1), (X_2, \Delta_2), \dots, (X_n, \Delta_n))$

Problème: estimer  $S_T$  de T

J'ai deux échantillons  $(X_1, \Delta_1), (X_2, \Delta_2), \dots, (X_n, \Delta_n)$  et

$((X'_1, \Delta'_1), (X'_2, \Delta'_2), \dots, (X'_n, \Delta'_n))$

Problème: comparer  $S_T$  et  $S_{T'}$

### ESTIMATEUR DE KAPLAN-MEIER

Si on a un échantillon  $(X_1, X_2, \dots, X_n)$  d'une loi de probabilité X, un estimateur de sa fonction de survie  $S_X(x) = P(X > x)$  est:

$$S_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{X_i > x}.$$

Cependant, il n'est pas souvent très aisé d'estimer  $S_X$  par  $S_n$  si on a des données censurées; ceci nous amène à définir l'estimateur suivant:

$$\widehat{S}_n(x) = \frac{1}{n} \left[ \sum_{i=1}^n 1_{X_i > x} + \sum_{i=1}^n \frac{\widehat{S}_n(x)}{\widehat{S}_n(X_i)} 1_{X_i \leq x} (1 - \Delta_i) \right]$$

On montre cette équation a une solution unique qui est:

$$(2.1) \quad \widehat{S}_n(x) = \prod_{X_i \leq x, \Delta_i=1} \left( 1 - \frac{1}{\sum_{j=1}^n 1_{(X_j \geq x_i)}} \right)$$

C'est l'estimateur de KAPLAN-MEIER

**Théorème 2.2.** Notons  $\tau_X = \inf\{x \geq 0 / S_X(x) = 0\}$

$\forall \tau < \tau_X$  on a:

$$1- \sup_{t \in [0, \tau]} |\widehat{S}_n(t) - S_T(t)| \xrightarrow[n \rightarrow \infty]{p.s.} 0$$

$$2- \forall t \in [0, \tau] \sqrt{n}(\widehat{S}_n(t) - S_T(t)) \xrightarrow[n \rightarrow \infty]{loi} N(0, V^2(t)) \text{ où } V^2(t) = -S_T^2(t) \int_0^t \frac{dS_T(u)}{S_T(u)S_C(u)}$$

On ne connaît pas  $V^2(t)$  mais un estimateur de  $V^2(t)$  est l'estimateur de GREENWOOD; il est convergent

## CONSTRUCTION DE L'ESTIMATEUR DE KAPLAN-MEIER

On ordonne les instants de décès distincts:

$$X_1 < X_2 < \dots < X_L$$

$$S_T(x_i) = P(T > x_i)$$

$$= P(T > x_i, T > x_{i-1})$$

$$= P(T > x_i / T > x_{i-1}) P(T > x_{i-1})$$

$$= P(T > x_i / T > x_{i-1}) P(T > x_{i-1} / T > x_{i-2}) \dots P(T > x_1 / T > x_0)$$

On suppose que l'individu a survécu au moins au delà de  $x_{i-1}$

On obtient:

$$P(T > x_i / T \geq x_i) \dots P(T > x_1 / T \geq x_1)$$

On estime  $P(T > x_i / T \geq x_i)$  par:

$$1 - \frac{1}{\sum_{j=1}^n 1_{(X_j \geq x_i)}}$$

d'où l'estimateur de KAPLAN-MEIER.

**Remarque 2.4.** En l'absence de censure, l'estimateur de KAPLAN-MEIER se ramène à  $S_n$  qui est l'estimateur de la fonction de survie

## 2.4 LE TEST DU LOG-RANK

### 2.4.1 PROCESSUS DE COMPTAGE

Soit un espace probabilisé  $(\Omega, F, P)$

Une famille de sous tribus  $(F_t, t \geq 0)$  de  $F$  est croissante si:  $\forall s \leq t$ , on a:  $F_s \subseteq F_t$

une telle famille est appelée filtration

**Définition 2.3.** Un processus stochastique est une famille de variables aléatoires  $(X_t)_{t \geq 0}$  définies sur  $(\Omega, F, P)$

**Définition 2.4.** à  $\omega$  dans  $\Omega$  fixé, la fonction  $t \mapsto X(\omega, t)$  s'appelle trajectoire du processus; ie à  $\omega$  fixé,  $X(\omega, t) = X_t(\omega)$

**Définition 2.5.** Un processus est dit continu à gauche si presque toutes ses trajectoires sont continues à gauche.

### EXEMPLE:Processus de comptage

C'est un processus stochastique  $(N_t, t \geq 0)$  continu à droite tel que  $N_0 = 0$  et  $N_t < \infty$  p.s

On a posé précédemment  $X = \min(X, T)$  pour nos données censurées

1- Soit  $N_t = 1_{X \leq t, \Delta=1} = 1_{X \leq t} \Delta$ ; si  $N_t = 1$  on a observé un décès avant t; le processus  $N_t$  compte le nombre de décès survenus avant t.

2-On définit le processus  $Y_t$  par:

$Y_t = 1_{X \geq t}$ .  $Y_t$  est un processus prévisible i.e sa valeur en t est déterminée juste avant t.

Les définitions de ces deux processus impliquent donc qu'il est équivalent de représenter une donnée censurée  $(X, \Delta)$  par  $(N_t, Y_t)$

## 2.4.2 LES MARTINGALES

Soit  $(\Omega, F, P)$  un espace probabilisé et  $(F_t, t \geq 0)$  une filtration de F.

Soit  $F_{t+} = \bigcap_{h>0} F_{t+h}$  et  $F_{t-} = \sigma(\bigcup_{h>0} F_{t-h})$  avec  $t > h$

Soit  $X = (X_t, t \geq 0)$  une famille de variables aléatoires réelles.

$X$  est une martingale si:

i-  $E[|X_t|] < \infty \forall t$

ii-  $X_t$  est  $F_{t-}$ -mesurable  $\forall t$  i.e le processus  $(X_t)_t$  est adapté à la filtration

iii-  $E[X_{t+s}/F_t] = X_t$  p.s  $\forall s, t \geq 0$

### PROPRIETES

$E[X_s/F_0] = X_0$  par definition

dès lors  $E[E[X_s/F_0]] = E[X_0]$  donc,  $\forall s$ , on a:

$E[X_s] = E[X_0]$

On voit clairement que l'espérance d'une martingale est constante.

### Exemple

Soient  $X_1, X_2, \dots, X_n$  n variables aléatoires indépendantes et identiquement distribuées (iid).

Soit  $S_n = \sum_{k=1}^n X_k$ : "marche aléatoire"

Soit  $F_n = \sigma(X_1, X_2, \dots, X_n)$

Il apparaît clairement que:

$E[S_{n+m}/F_n] = S_n$ ; donc, la "marche aléatoire" est une martingale.

Pour la définition de la martingale, on avait:  $E[X_{t+s}/F_t] = X_t$ ;

si maintenant on a:

$E[X_{t+s}/F_t] \geq X_t$ , on dit que  $X_t$  est une sous-martingale.

Si  $E[X_{s+t}/F_t] \leq X_t$  on dit que  $X_t$  est une surmartingale

## EXEMPLES

1- Une martingale est à la fois une sous-martingale et une sur-martingale.

2- si  $\varphi : R \rightarrow R$  convexe et  $X_t$  est une martingale, alors:  $\varphi \circ X_t$  est une sous-martingale.

En effet,

$E[\varphi \circ X_{t+s}/F_t] \geq \varphi \circ E[X_{t+s}/F_t] = \varphi \circ X_t$  ceci, d'après l'inégalité de Jensen.

## DECOMPOSITION DE DOOB-MEYER

**Théorème 2.3.** Soit  $X$  une sous-martingale positive.

Alors, il existe une martingale  $M$  et un processus croissant prévisible  $A$  tel que:

$$A_0 = 0 \text{ et } X_t = A_t + M_t$$

**Remarque 2.5.** Cette décomposition est unique

On montre que:  $A_t = \int_0^t Y(u) \lambda_T(u) du$  où  $Y(x) = 1_{X \geq x}$

**Remarque 2.6.** Soit  $X_t$  une martingale. Alors:

$$1- E[X_t/F_{t-h}] = X_{t-h}$$

$$2- E[X_t - X_{t-h}/F_{t-h}] = 0$$

$$3- E[X_t - X_{t-}/F_{t-}]$$

## L'estimateur de Nelson-Aalen

$N_t$  est une sous-martingale positive; donc d'après la décomposition de DOOB-MEYER,:

$N_t = A_t + Z_t$  où  $A_t = \int_0^t Y(u) \lambda_T(u) du$  est un processus croissant prévisible et  $Z_t$  une martingale.

On a:

$$\begin{aligned} Z_t &= N_t - A_t \\ &= \int_0^t dN_s - \int_0^t Y(u) \lambda(u) \\ &= \int_0^t (dN_s - Y(s) \lambda(s)) \end{aligned}$$

ainsi,

$dZ_t = dN_t - Y(t) \lambda(t) dt$  or  $Z_t$  est une martingale, donc d'après la remarque précédente, on

a:

$$E[dZ_t/F_{t-}] = 0$$

d'où:

$$E[dN_t/F_{t-}] = E[Y(t)\lambda(t)dt/F_{t-}]$$

or à l'instant t,  $\lambda(t) = cste$  donc

$$E[dN_t/F_{t-}] = \lambda(t)E[Y(t)dt/F_{t-}];$$

mais aussi, on sait que  $Y_t = 1_{Y \geq t}$  est un processus continu à gauche, donc  $F_{t-}$ -mesurable; dès lors:

$$E[dN_t/F_{t-}] = \lambda(t)Y(t)dt$$

i.e:

$$\lambda(t)dt = \frac{E[dN_t/F_{t-}]}{Y(t)}$$

On avait défini  $\Lambda(u) = \int_0^u \lambda(t)dt$ , donc:

$$\Lambda(u) = \int_0^u \frac{E[dN_t/F_{t-}]}{Y(t)}.$$

Nous allons estimer  $\Lambda$ ; pour cela, il nous faut estimer l'espérance.

De façon générale, on note:  $P_n(f(X)) = \frac{1}{n} \sum_{i=1}^n f(X_i)$  l'estimateur de  $f(X)$

On a:

$$E[dN_t] = \lambda(t)dtE[Y(t)]$$

i.e

$$P_n(dN_t) = P_n(Y(t)\lambda(t)dt)$$

$$= \frac{1}{n} \sum_{i=1}^n Y_i(t)\lambda(t)dt$$

$$= \lambda(t)dtP_n(Y(t))$$

d'où

$$dt\hat{\lambda}(t) = \frac{P_n(dN_t)}{P_n(Y(t))};$$

il vient donc que:

$$\hat{\Lambda}(t) = \int_0^t \frac{P_n(dN_u)}{P_n(Y(u))}$$

$$= \sum_{k:t_k \leq t} \frac{\text{nombre de censored } k}{\text{nombre d'individus risquant } k}:$$

$\hat{\Lambda}$  est l'estimateur de Nelson-Aalen de  $\Lambda$

## PROPRIETES DE L'ESTIMATEUR DE NELSON-AALEN

**Théorème 2.4.** Soit  $\tau = \sup_t \{P(X > t) > 0\}$

$\forall t < \tau$  on a:

$$1 - \sup_{s \in [0, t]} |\hat{\Lambda}_{NA}(s) - \Lambda(s)| \xrightarrow{p.s} 0$$

$$2 - \sqrt{n}(\hat{\Lambda}_{NA}(t) - \Lambda(t)) \xrightarrow[n \rightarrow \infty]{loi} N(0, V^2(t))$$

### Le Test Du Log-rank

On a une population de n individus répartie en k groupes; On va tester

$H_0 : \lambda_1 = \lambda_2 = \dots = \lambda_k$  contre:

$H_1 : \exists i, j \text{ dans } [1, k] / \lambda_i \neq \lambda_j$

Le principe de test reste le même que celui dans 2 groupes que nous allons considérer ici pour

simplifier.

Soit donc à tester:

$$H_0 : \lambda_1 = \lambda_2 \text{ contre}$$

$$H_1 : \lambda_1 \neq \lambda_2$$

L'idée du test est de considérer qu'effectuer le test précédent est équivalent à tester:

$$H_0 : \Lambda_1 = \Lambda_2 \text{ contre}$$

$$H_1 : \Lambda_1 \neq \Lambda_2$$

## NOTATIONS

On a 2 groupes, L instants de décès distincts.

$\overline{\Delta N}(t_k)$  = nombre de décès en  $t_k$  dans le groupe  $i$

$$\overline{\Delta N}(t_k) = \overline{\Delta N}_1(t_k) + \overline{\Delta N}_2(t_k)$$

$\overline{Y}_i(t_k)$  = nombre d'individus à risque en  $t_k$  dans le groupe  $i$

$\overline{Y}(t_k)$  = nombre de sujets à risque en  $t_k$

$$\widehat{\Lambda}_{NA}(t) = \sum_{t_k < t} \frac{\overline{\Delta N}(t_k)}{\overline{Y}(t_k)}$$

$$\widehat{\Delta}_{NA,i}(t) = \sum_{t_k \leq t} \frac{\overline{\Delta N}_i(t_k)}{\overline{Y}_i(t_k)}$$

La statistique de test est:

$$U_i = \sum_{k=1}^L \overline{Y}_i(t_k) \left[ \frac{\overline{\Delta N}_i(t_k)}{\overline{Y}_i(t_k)} - \frac{\overline{\Delta N}(t_k)}{\overline{Y}(t_k)} \right] \quad i=1 \text{ ou } i=2$$

## REGLE DE DECISION

On rejette  $H_0$  si:

$$1 - \left| \frac{U_i}{\sqrt{\text{var}(U_i)}} \right| > U_{1-\frac{\alpha}{2}} \text{ de } N(0,1)$$

ou si:

$$2 - \frac{U_i^2}{\text{var}(U_i)} > U_{1-\alpha} \text{ du khi deux à un degré de liberté.}$$

# APPLICATIONS DES METHODES ET RESULTATS

---

## 3.1 INTRODUCTION

La résolution du problème qui nous était posé était conditionnée par la disponibilité des données complètes. Pour analyser les données recueillies, les temps d'envol des moustiques ont été regroupés par intervalles de temps; le principe était de compter le nombre de moustiques envolés par intervalle de temps; la subdivision a été faite de la manière suivante:

[0,1[; [1,2[; [2,4[; [4,8[; [8,16[; [16,32[; [32,64[; [64,128[; [128,256]

Ce nombre de moustiques envolés par intervalles de temps ne nous permet pas d'effectuer notre analyse de survie; mais seulement, nous pouvons avoir pour chaque insecticide et pour chaque mois la fonction de répartition empirique dont ces réalisations sont issues. Nous allons donc simuler ces données comme la méthode du Chapitre II l'a décrit; pour simuler ces données nous avons une indication: les temps d'envol des moustiques lors du premier mois pour les quatre insecticides sont complets; nous commencerons donc par rechercher la loi dont ces réalisations sont issues, nous simulerons ensuite les données conformément à l'extrapolation faite et aux méthodes exposées.

## 3.2 RECHERCHE DE LA LOI

La recherche de la loi a été motivée par l'existence dans la littérature de certaines lois qui régissent souvent les phénomènes étudiés en analyse de survie

### LA LOI LOG-NORMALE

Pour savoir si les temps d'envol étaient issus d'une certaine loi de probabilité, nous avons utilisés des tests statistiques dont la théorie n'a pas été mentionnée ici, puisque le problème principal n'y faisait pas directement appel. Le test que nous avons utilisé ici est celui de Shapiro.

Insecticides	permethrine	cyfluthrine	deltasolution	deltatablettes
Pvalue	1.44e-9	0.00019	0.0134	2.619e-5

Aucune des p-value n'est supérieure à 0.05, donc nous rejettons l'hypothèse qui nous dit que ces temps d'envol sont issus d'une loi log-normale.

### LA LOI GAMMA

Comme nous l'avons dit au chapitre II, la loi Gamma ne fait pas partie des lois dont l'estimateur du maximum de vraisemblance n'est déterminable que numériquement par des approximations successives. Nous avons donc implémenté la procédure qui estime les paramètres d'une loi Gamma à base de n réalisations observées, l'initialisation étant faite par la méthode des moments comme nous l'avons annoncé au Chapitre II; le code R de ce programme est donné en annexe du document. Nous avons obtenu les estimations suivantes:

Insecticides	Permethrine	cyfluthrine	deltasolution	deltatable
a	-21090.576	0.561	0.6178	0.4630
lamda	23535.77	45.572	60.824	104.60

Pour savoir si nos données suivaient une loi Gamma, nous avons utilisé le test de Kolmogorov-Smirnov; les différentes p-values étaient:

Insecticides	permethrine	cyfluthrine	deltasol	deltatable
pvalue	nan	2.21e-16	2.21e-16	2.2e-16

Ce tableau de p-values nous montre que nos données ne suivent pas une loi de Gamma.

### LA LOI DE WEIBULL

L'estimateur du maximum de vraisemblance de la loi ne pouvant pas être déterminés de façon explicite, nous avons implémenté la procédure qui estime les paramètres de la loi de weibull par le maximum de vraisemblance, l'initialisation étant faite de façon aléatoire. Nous avons obtenu les paramètres suivants:

Insecticides	permethrine	cyfluthrine	deltasol	deltatable
alpha	1.422	0.681	0.7106	41.096
lambda	2.781	19.813	28.806	16.629

Comme avec la loi Gamma, nous effectuons le test de Kolmogorov-Smirnov pour savoir si nos données suivent une loi de Weibull; les résultats obtenus sont:

Insecticides	permethrine	cyfluthrine	deltasol	deltatable
pvalue	0.71	0.49	0.89	0.54

### 3.3 ESTIMATION DES AUTRES PARAMETRES

Puisque nous avons déterminé notre loi qui est celle de Weibull, nous simulons les temps d'envol des moustiques des autres mois suivant une loi de Weibull; il se pose un problème: quels sont les paramètres pour chaque mois et pour chaque loi qui nous permettront de simuler les données voulues? Nous allons donc appliquer la méthode d'estimation annoncée au II-1-4; Nous nous contentons de décrire l'algorithme ici; la procédure est présentée en annexe du document.

#### ALGORITHME

- a- Calcul de la distance entre les deux fonctions
    - lire les deux vecteurs  $n$  et  $x$ , représentant respectivement le nombre de moustiques envolés par intervalles de temps et le vecteur ayant servi à la subdivision.
    - si l'un des paramètres est négatif ou nul, associer à la distance la valeur 20.
    - Sinon, calculer la distance entre les deux fonctions.
  - b-initialiser les paramètres par une technique d'initialisation aléatoire, en tenant compte des signes des paramètres.
  - c-Utiliser la fonction  $nlm$  de R en s'assurant au préalable que les conditions pour obtenir un minimum sont vérifiées.
  - d- Faire tourner un grand nombre de fois pour avoir non un minimum local mais un minimum global.
- La procédure R de l'algorithme est fournie en annexe du mémoire.

### 3.4 PRESENTATION DES RESULTATS DE L'ANALYSE DE SURVIE

#### MOIS I

- A- OPPOSITION DES QUATRE INSECTICIDES
  - 1- COURBES DE KAPLAN-MEIER

### Courbes de Survie des quatre insecticides

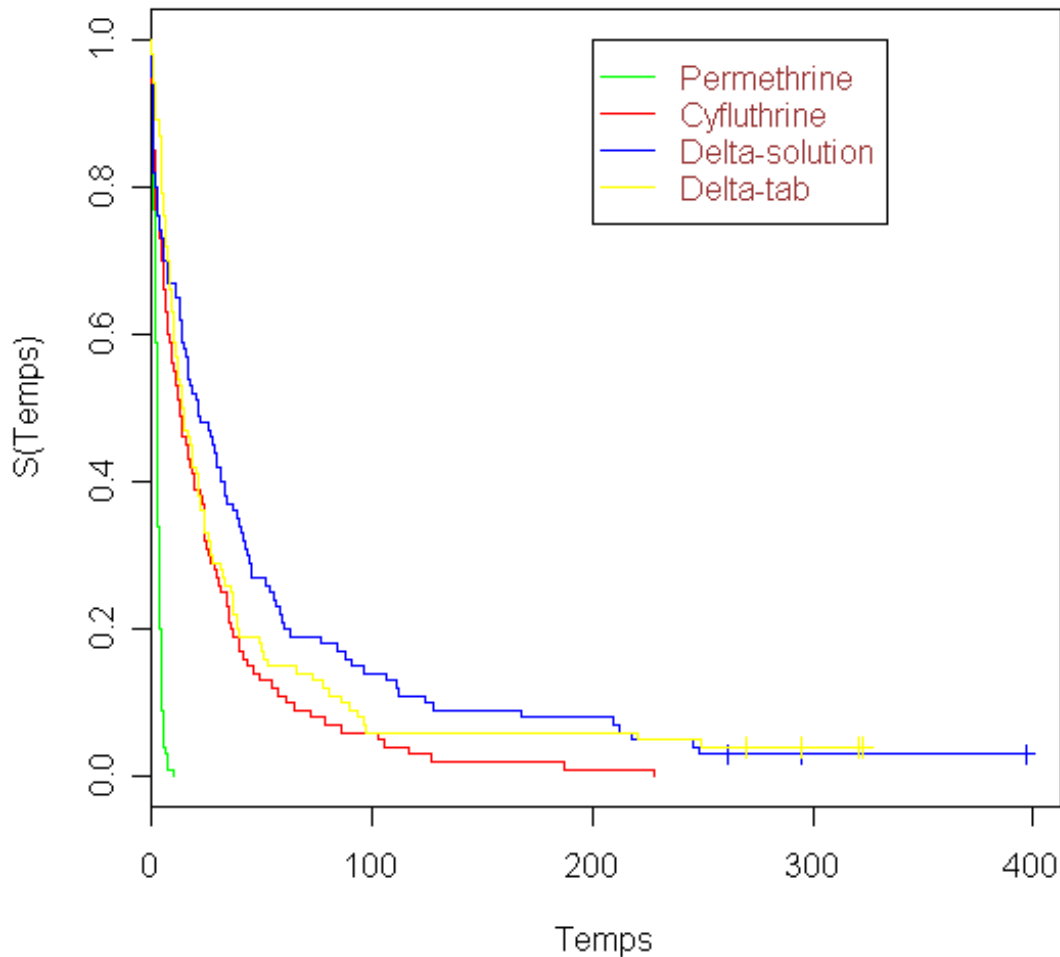


FIG. 3.1 – *fig1*

Les courbes de kaplan-meier nous donnent une idée sur le comportement de l'irritabilité de nos quatre insecticides ; ainsi, on peut à priori penser que la permethrine pourrait être meilleure que les autres insecticides, ce qui nous conduit à effectuer le test du log-rank, pour affirmer ou infirmer l'hypothèse à priori que nous fournit les courbes.

#### 2- TEST DU LOG-RANK

```
Call: survdiff(formula = obj1  insect, rho = 1)
      N Observed Expected (O - E)2/E(O - E)2/V
insect=1 100 74.6 25.7 92.99 140.5
insect=2 100 46.2 54.1 1.16 2.4
insect=3 100 40.0 60.2 6.80 15.4
insect=4 100 40.0 60.7 7.09 15.5
      Chisq= 143 on 3 degrees of freedom, p= 0
```

La P-value de notre test est nulle; ce qui nous conduit à rejeter l'hypothese d'egalite des courbes de survie; d'où on peut dire que les quatre insecticides ne sont pas irritants de la même façon ;il convient donc de savoir lequel est le plus irritant ou le moins irritant des quatre insecticides

B-OPPOSITION DE LA CYFLUTHRINE ET DE LA DELTAMETHRINE EN SOLUTION  
 1-COURBES DE KAPLAN-MEIER On peut penser d'après les courbes de KAPLAN-

**ourbes de survie de la cyfluthrine et de la deltamethrine\_sol**

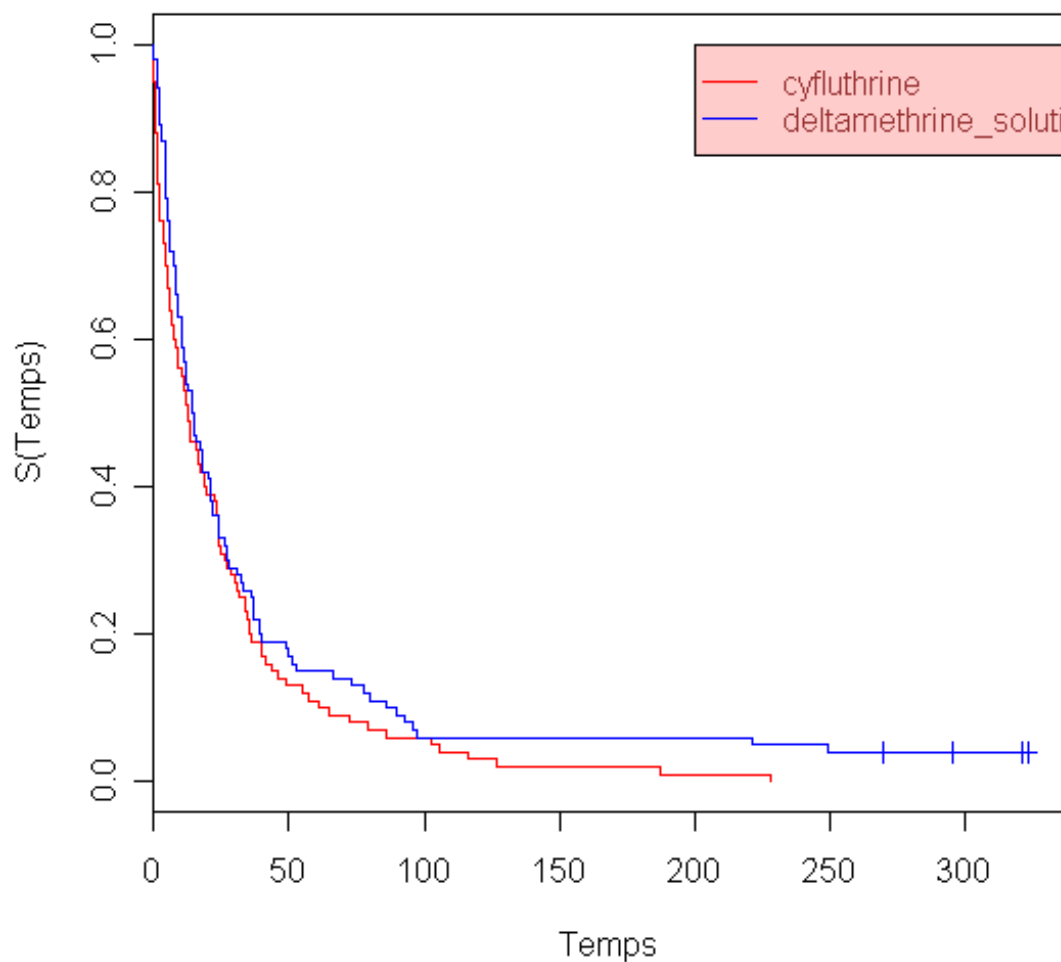


FIG. 3.2 – fig2

MEIER que les deux insecticides ont le même effet d'irritabilité.

2-TEST DU LOG-RANK

Call: survdiff(formula = obj1 insect, rho = 0)

N Observed Expected (O - E)<sup>2</sup>/E(O - E)<sup>2</sup>/V  
 insect=1 100 100 90 1.098 2.07

insect=2 100 96 106 0.933 2.07

Chisq= 2.1 on 1 degrees of freedom, p= 0.15

La p-value etant de 0.15, il vient que la cyfluthrine et la deltamethrine en solution on le même effet d'irritabilité lors du premier mois

C- OPPOSITION DE LA DELTAMETHRINE EN TABLETTES ET DE LA DELTAMETHRINE EN SOLUTION

1-COURBES DE KAPLAN-MEIER

### courbes de survie des deux deltamethrines

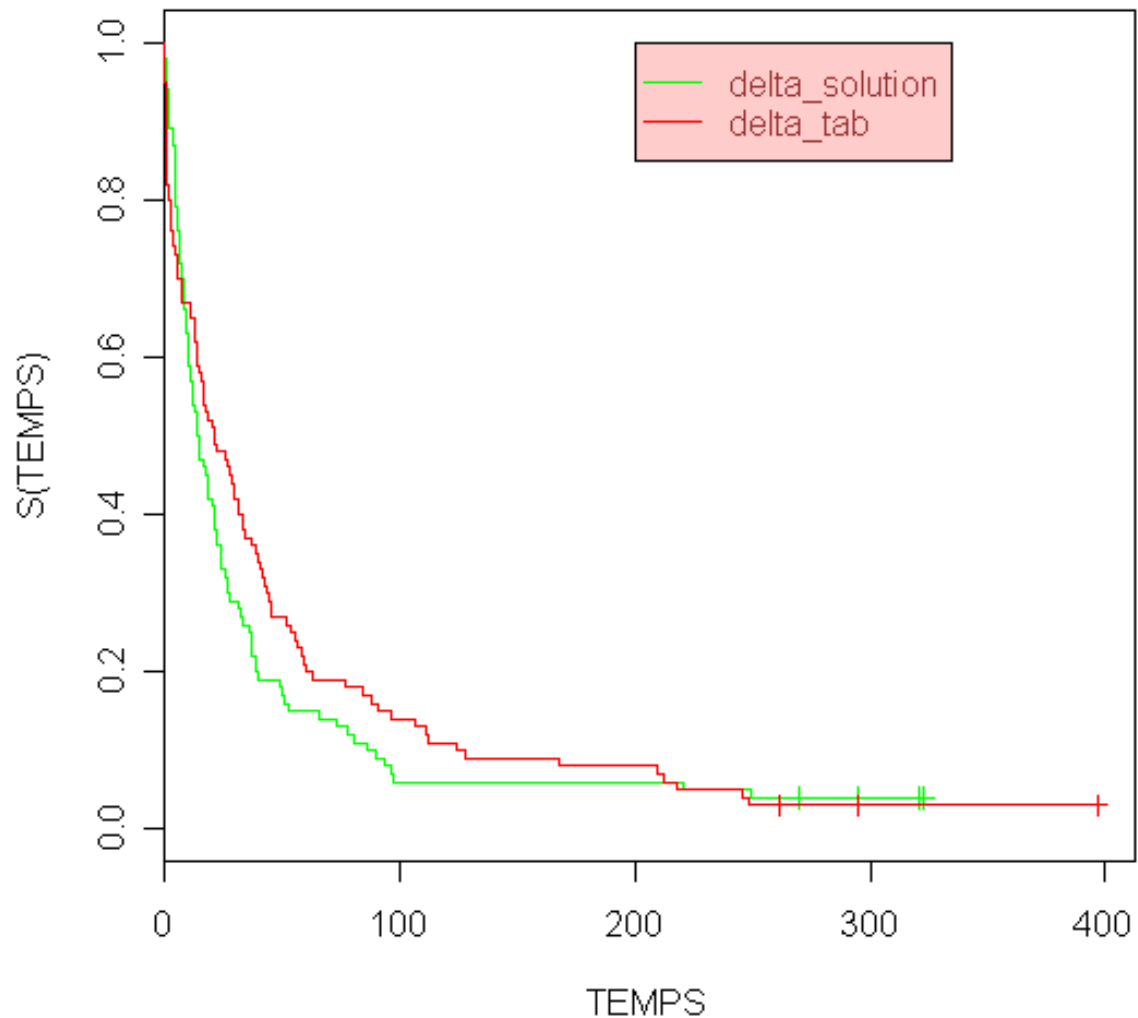


FIG. 3.3 – *fig3*

A priori, les courbes de kaplan meier nous montrent que les deux insecticides n'ont pas véritablement le même effet d'irritabilité, la différence n'étant pas toutefois significative.

2-TEST DU LOG-RANK

Call: survdiff(formula = obj1 insect, rho = 0)

	N	Observed	Expected	$(O - E)^2/E$	$(O - E)^2/V$
insect=1	100	96	88.7	0.598	1.13
insect=2	100	97	104.3	0.509	1.13

Chisq= 1.1 on 1 degrees of freedom, p= 0.287

La valeur de la p-value nous montre que la difference d' effet d'irritabilié entre les deux types de deltamethrine n'est pas significative:les deux insecticides ont le même effet.

On peut donc dire en se basant sur les résultats précédents que la permethrine lors du premier mois est meilleure que les trois autres insecticides, qui ont le même effet d'irritabilité.

**MOIS II**

A CONFRONTATION DES QUATRE INSECTICIDES

1- COURBES DE KAPLAN-MEIER Les courbes de survie ci-dessus nous proposent que la

**courbes de survie des quatre insecticides**

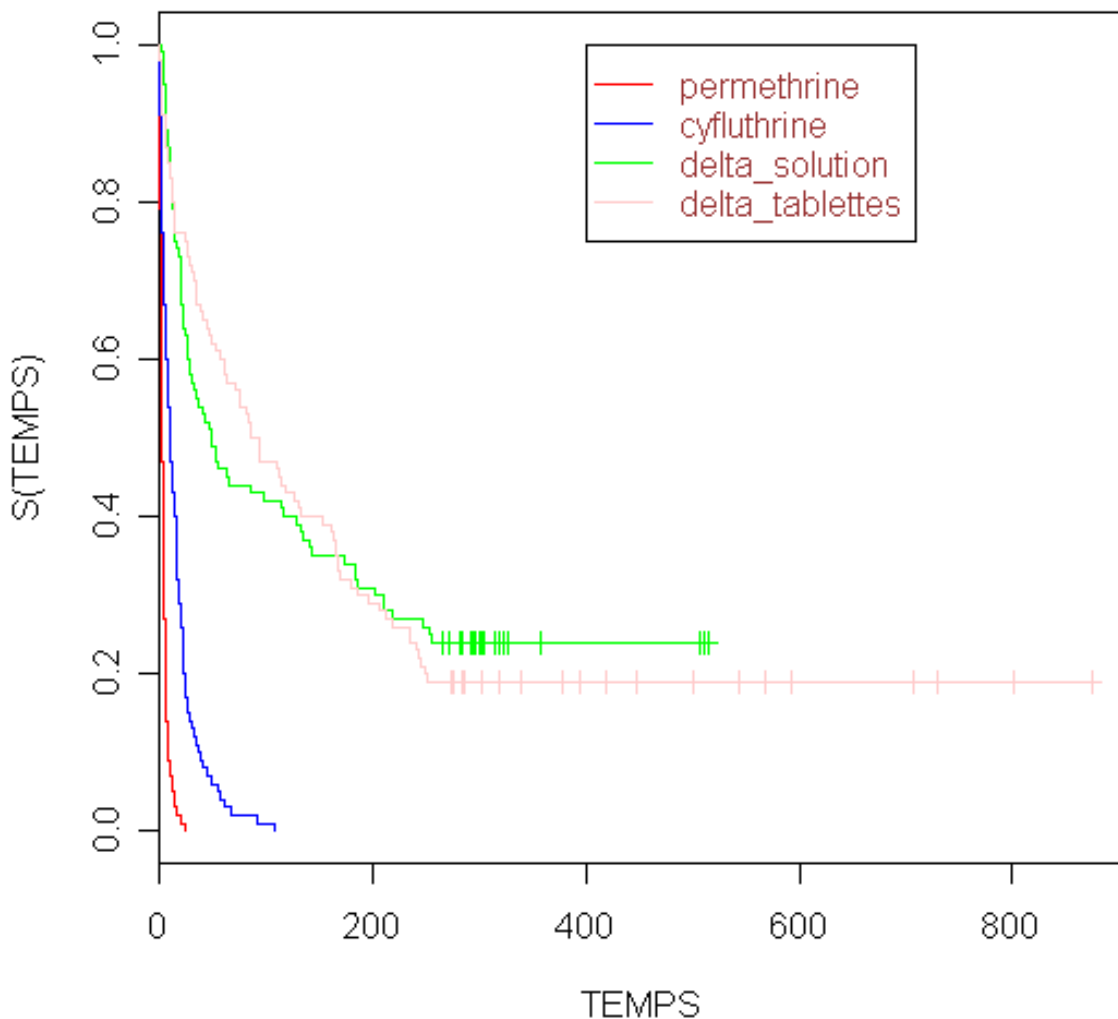


FIG. 3.4 – fig4

permethrine et la cyfluthrine pourraient avoir le même effet, tandis que les deux deltaméthrines pourraient de leur côté avoir le même effet d'irritabilité.

## 2- LE TEST DU LOG-RANK

Les résultats du test sont les suivants :

Call: survdiff(formula= obj1 insect, rho = 0)

N Observed Expected  $(O - E)^2/E$   $(O - E)^2/V$

insect=1 100 100 25.9 211.5 253.9

insect=2 100 100 59.3 27.8 35.8

insect=3 100 76 133.3 24.6 40.3

insect=4 100 81 138.4 23.8 40.5

Chisq= 342 on 3 degrees of freedom, p= 0

La p-value de notre test est numme, ce qui nous conduit à rejeter l'hypothèse nulle, donc, les quatre insecticides n'ont pas le même effet d'irritabilité.

## B-OPPOSITION ENTRE LA PERMETHRINE ET LA CYFLUTHRINE

### 1- COURBES DE KAPLAN-MEIER

**courbes de survie de la permethrine et de la cyfluthrine**

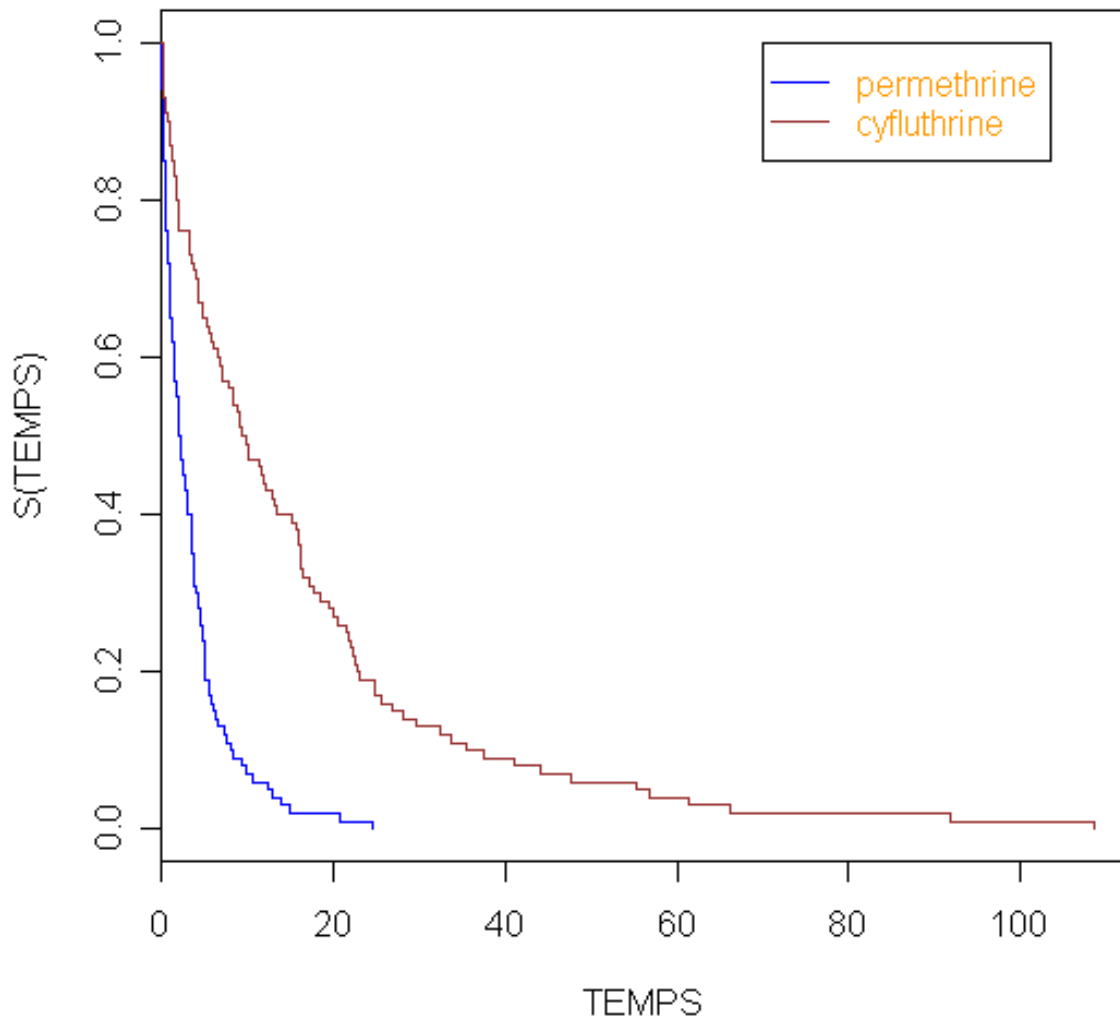


FIG. 3.5 – fig5

Les dimensions des temps étant très petites par rapport aux valeurs maximales des deux autres insecticides, on voit ici que la permethrine et lacyfluthrine n’ont pas le même effet d’irritabilité, comme peut le laisser croire le schéma précédent.

2- LE TEST DU LOG-RANK

Call: survdiff(formula = obj1 insect, rho = 0)

N Observed Expected  $(O - E)^2/E(O - E)^2/V$  insect=1 100 100 54.3 38.4 61.6  
 insect=2 100 100 145.7 14.3 61.6

Chisq= 61.6 on 1 degrees of freedom, p= 4.22e-15 La p-value de notre test est prati-

quement nulle, donc on rejette l'hypothese d'egalité des courbes de survie; ainsi, nos deux insecticides n'ont pas le meme degré d'irritabilité. La permethrine est donc plus irritante que la cyfluthrine.

C- OPPOSITION DES DEUX DELTAMETHRINES

1- COURBES DE KAPLAN-MEIER

**courbes de survie des deux deltamethrines**

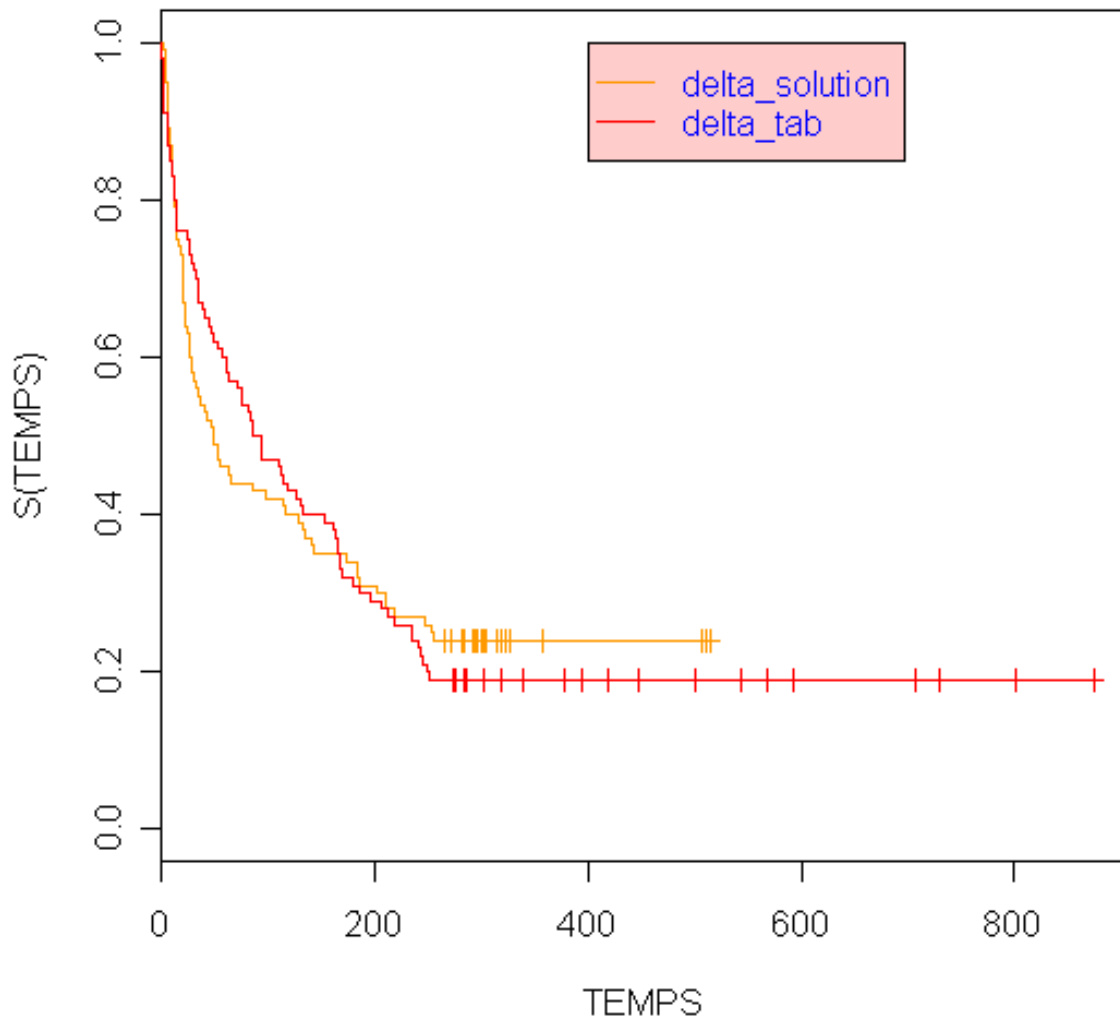


FIG. 3.6 – *fig6*

2- LE TEST DU LOG-RANK Call: survdiff(formula = obj1 insect, rho = 0)

N Observed Expected  $(O - E)^2/E$   $(O - E)^2/V$  insect=1 100 76 76.4 0.00167 0.00328

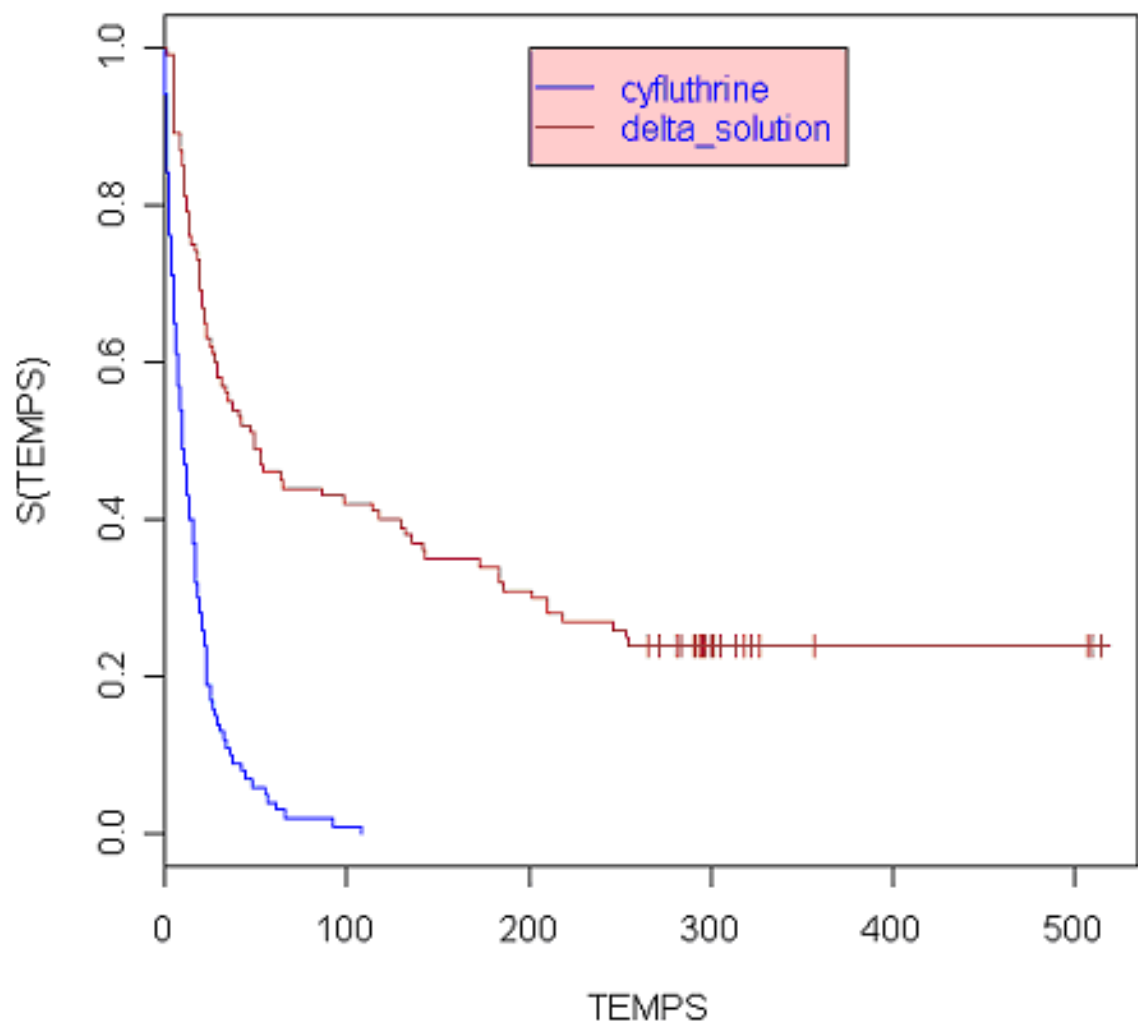
insect=2 100 81 80.6 0.00158 0.00328

Chisq= 0 on 1 degrees of freedom, p= 0.954

La p-value de notre test nous informe qu'il n'y a pas de différence significative d'effet d'irritabilité entre les deux deltaméthrines.

D-OPPOSITION ENTRE LA CYFLUTHRINE ET LA DELTAMETHRINE EN SOLUTION

1- COURBES DE KAPLAN-MEIER

**courbes de survie de la cyfluthrine et de la delta\_solutior**FIG. 3.7 – *fig7*

2- TEST DU LOG-RANK Call: survdiff(formula = obj1 insect, rho =0)

N Observed Expected  $(O - E)^2/E$   $(O - E)^2/V$

insect=1 100 100 50.3 49.1 77.8

insect=2 100 76 125.7 19.7 77.8

Chisq= 77.8 on 1 degrees of freedom, p= 0

La p-value de notre test est nulle, donc, on peut conclure que la cyfluthrine est plus irritante que la deltamethrine en solution.

Au terme de cette analyse effectuée au second mois, il vient que la permethrine est l'insecticide le plus irritant, suivi de la cyfluthrine; les deux deltamethrines ont un effet d'irritabilité égal.

### MOIS III

A-OPPOSITION DES QUATRE INSECTICIDES

1-COURBES DE KAPLAN-MEIER

### Courbes de survie des quatre insecticides

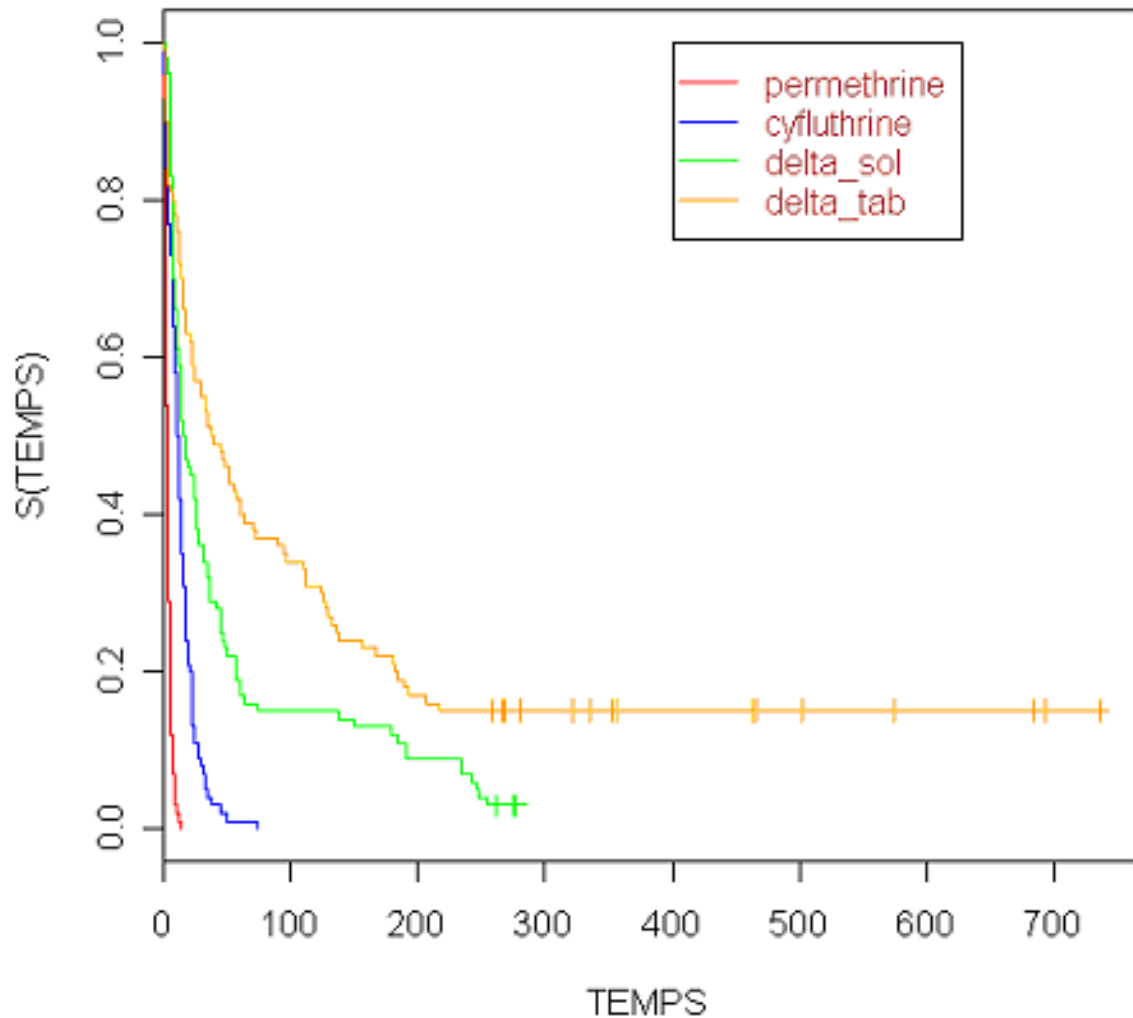


FIG. 3.8 – fig7

#### 2-TEST DU LOG-RANK

Call: survdiff(formula = obj1 insect, rho = 0)

	N	Observed	Expected	$(O - E)^2/E$	$(O - E)^2/V$
insect=1	100	100	25.7	214.98	259.4
insect=2	100	100	75.4	8.02	10.7
insect=3	100	97	123.7	5.77	8.6
insect=4	100	85	157.2	33.15	61.8

Chisq= 310 on 3 degrees of freedom, p= 0

La p-value de notre test est nulle donc, il ya une difference significative entre les effets d'irritabilité des quatres insecticides.

B- OPPOSITION DE LA CYFLUTHRINE ET DE LA PERMETHRINE  
 1- COURBES DE KAPLAN-MEIER

**courbes de survie de la permethrine et de la cyfluthrine**

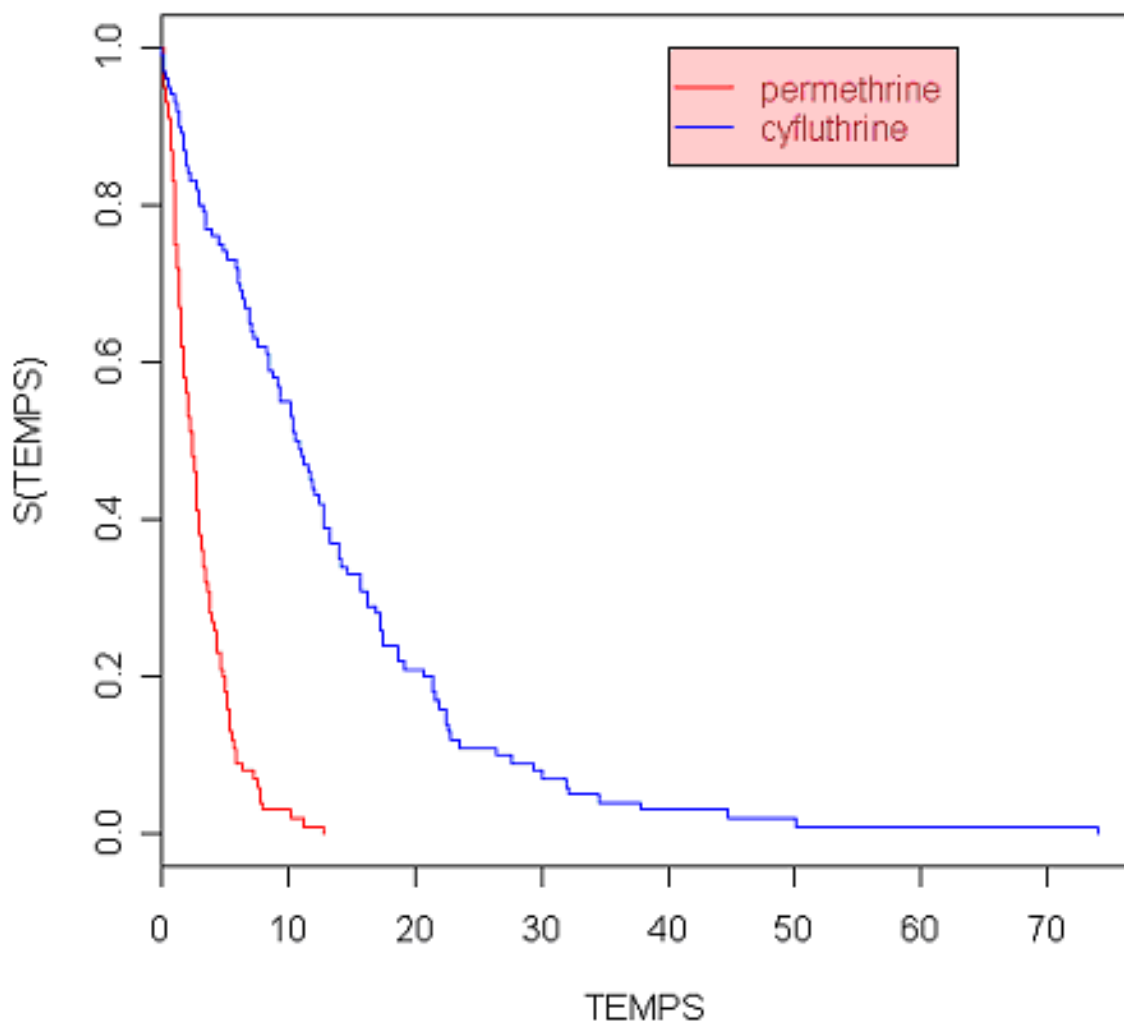


FIG. 3.9 – fig8

2- TEST DU LOG-RANK Call: survdiff(formula = obj1 insect, rho =0)

N Observed Expected  $(O - E)^2/E(O - E)^2/V$

insect=1 100 100 46.4 61.9 100

insect=2 100 100 153.6 18.7 100

Chisq= 100 on 1 degrees of freedom, p=0

La p-value de notre test est nulle, donc, on peut dire que la permethrine est meilleure que la cyfluthrine en terme d'irritabilité.

C-OPPOSITION DE LA CYFLUTHRINE ET DE LA DELTAMETHRINE EN SOLUTION  
1- COURBES DE KAPLAN MEIER

**courbes de survie de la cyfluthrine et de la delta\_ solution**

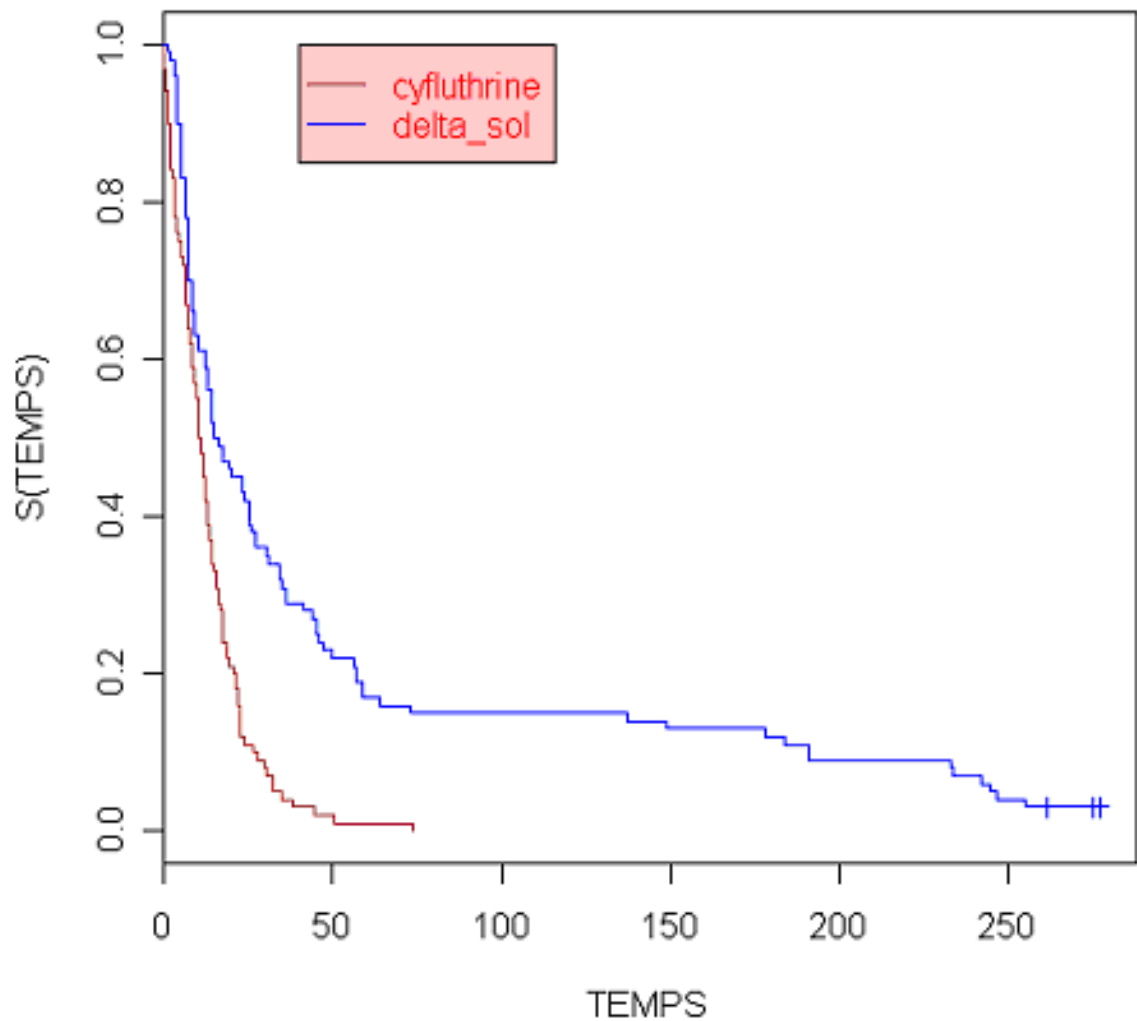


FIG. 3.10 – *fig9*

2- TEST DU LOG-RANK

Call: survdiff(formula = obj1 insect, rho = 0)

N Observed Expected  $(O - E)^2/E$   $(O - E)^2/V$

insect=1 100 100 67.1 16.14 27.9

insect=2 100 97 129.9 8.34 27.9

Chisq= 27.9 on 1 degrees of freedom, p= 1.3e-07 La p-value de notre test est quasiment nulle, donc la cyfluthrine irrite plus que la deltamethrine en solution.

D-OPPOSITION DES DEUX DELTAMETHRINES

1- COURBES DE KAPLAN-MEIER 2- TEST DU LOG- RANK

**courbes de survie des deux deltamethrines**

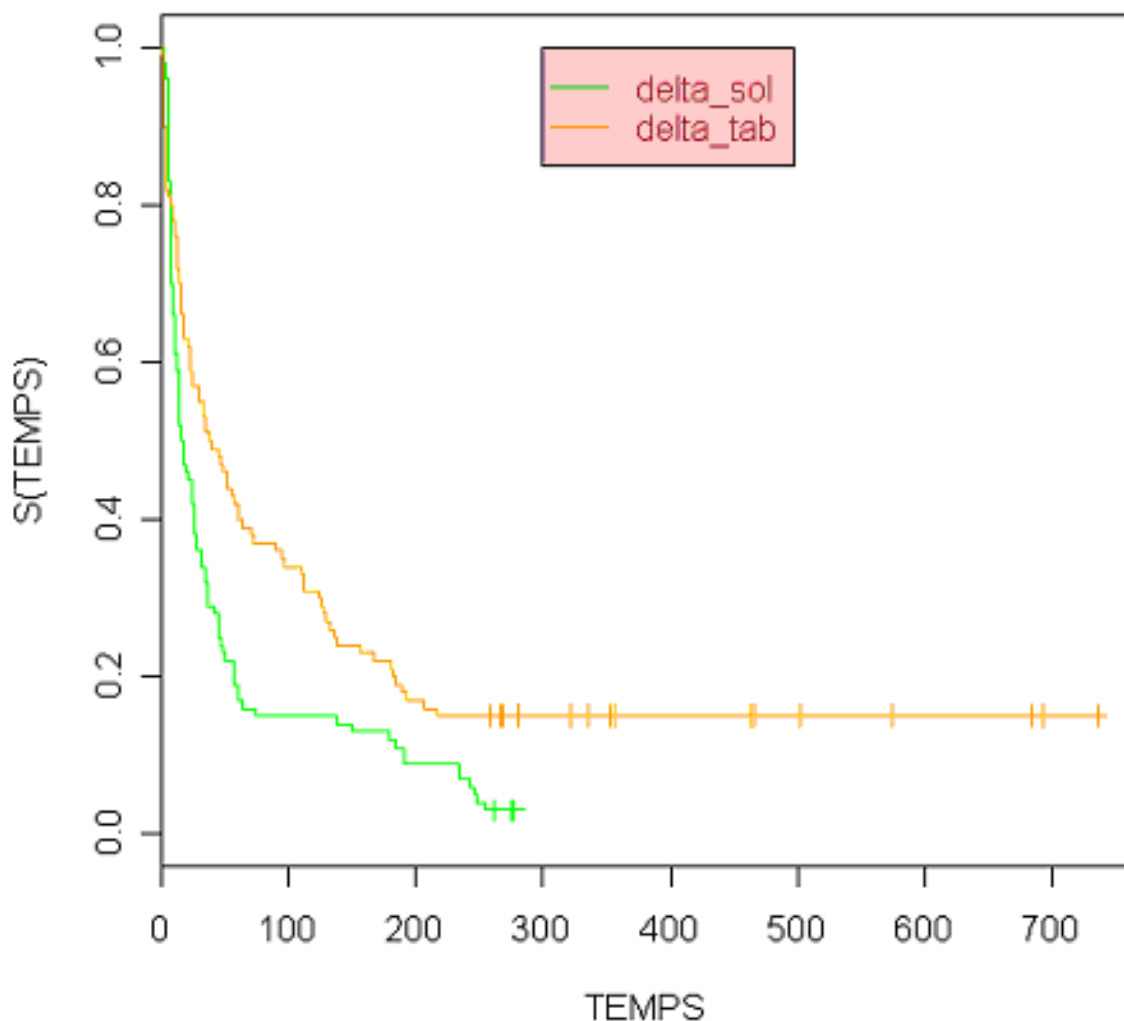


FIG. 3.11 – fig10

Call: survdiff(formula = obj1 insect, rho = 0)

N Observed Expected  $(O - E)^2/E$   $(O - E)^2/V$

insect=1 100 97 76.6 5.42 9.67

insect=2 100 85 105.4 3.94 9.67

Chisq= 9.7 on 1 degrees of freedom, p= 0.00187 La p-value de notre test est de  $0.00187 < 0.05$ , donc dans les conditions de l'expérience, on peut dire que la deltaméthrine en solution est meilleure que la deltaméthrine en tablettes.

L'analyse de survie nous montre que la permethrine lors du troisième mois est l'insecticide le plus irritant, suivi de la cyfluthrine qui est plus irritante que la deltaméthrine en solution, qui elle même est plus irritante que la deltaméthrine en tablettes.

**MOIS IV**

A-OPPOSITION DES QUATRE INSECTICIDES  
 1-COURBES DE KAPLAN-MEIER 2- Le TEST DU LOG-RANK

**COURBES DE SURVIE DES QUATRE INSECTICIDES**

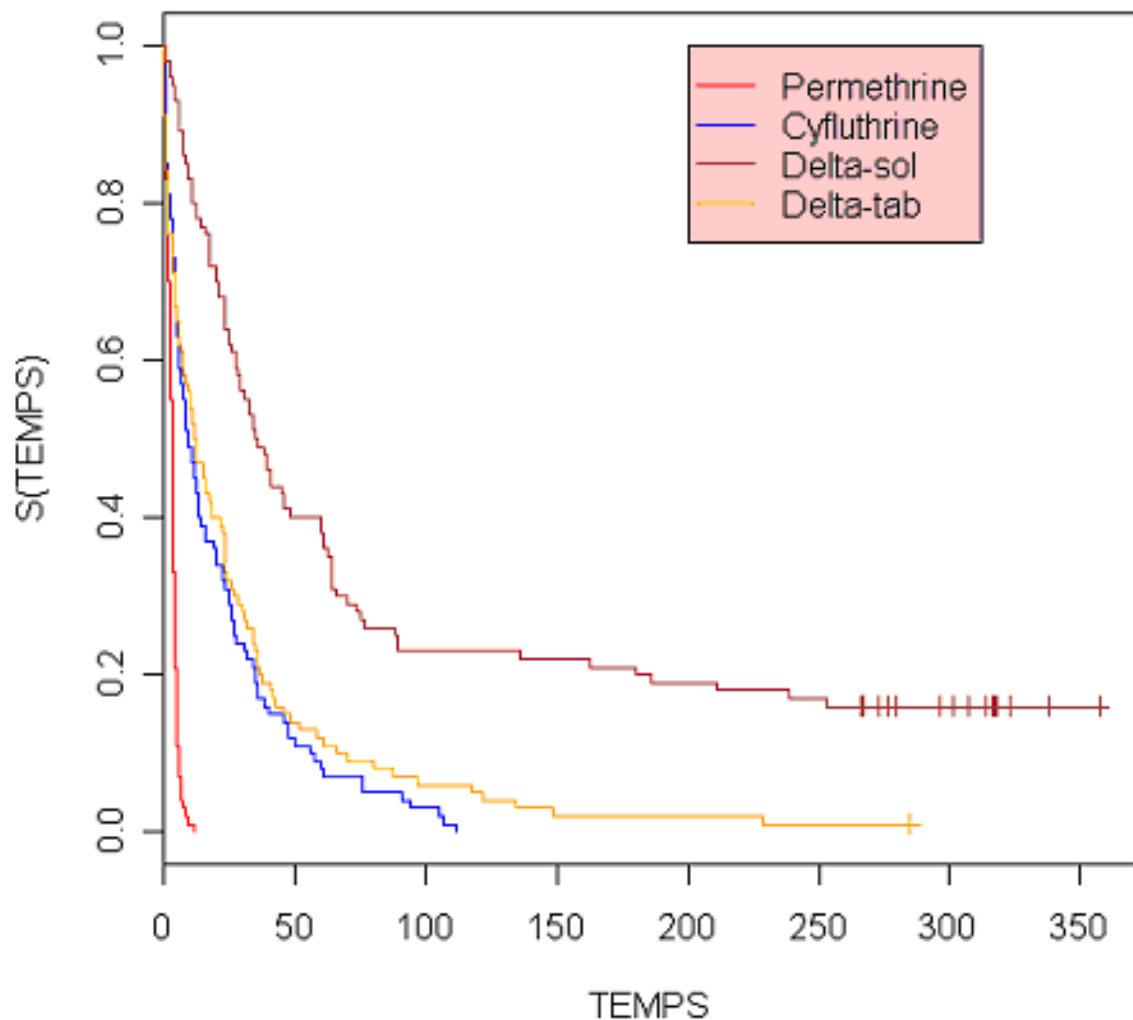


FIG. 3.12 – fig11

Call: survdiff(formula = obj1 hospi, rho = 0)  
 N Observed Expected  $(O - E)^2/E$   $(O - E)^2/V$

hospi=1 100 100 31.6 148.178 184.493  
 hospi=2 100 100 87.8 1.709 2.264  
 hospi=3 100 84 168.3 42.211 82.569  
 hospi=4 100 99 95.4 0.137 0.185

Chisq= 227 on 3 degrees of freedom, p= 0

La p-value de notre test est nulle, donc les insecticides ont des effets d'irritabilité significativement différents.

B-OPPOSITION DE LA PERMETHRINE ET DE LA CYFLUTHRINE  
 1-COURBES DE KAPLAN-MEIER 2- TEST DU LOG-RANK

**OPPOSITION DE LA PERMETHRINE ET DE LA CYFLUTHRINE**

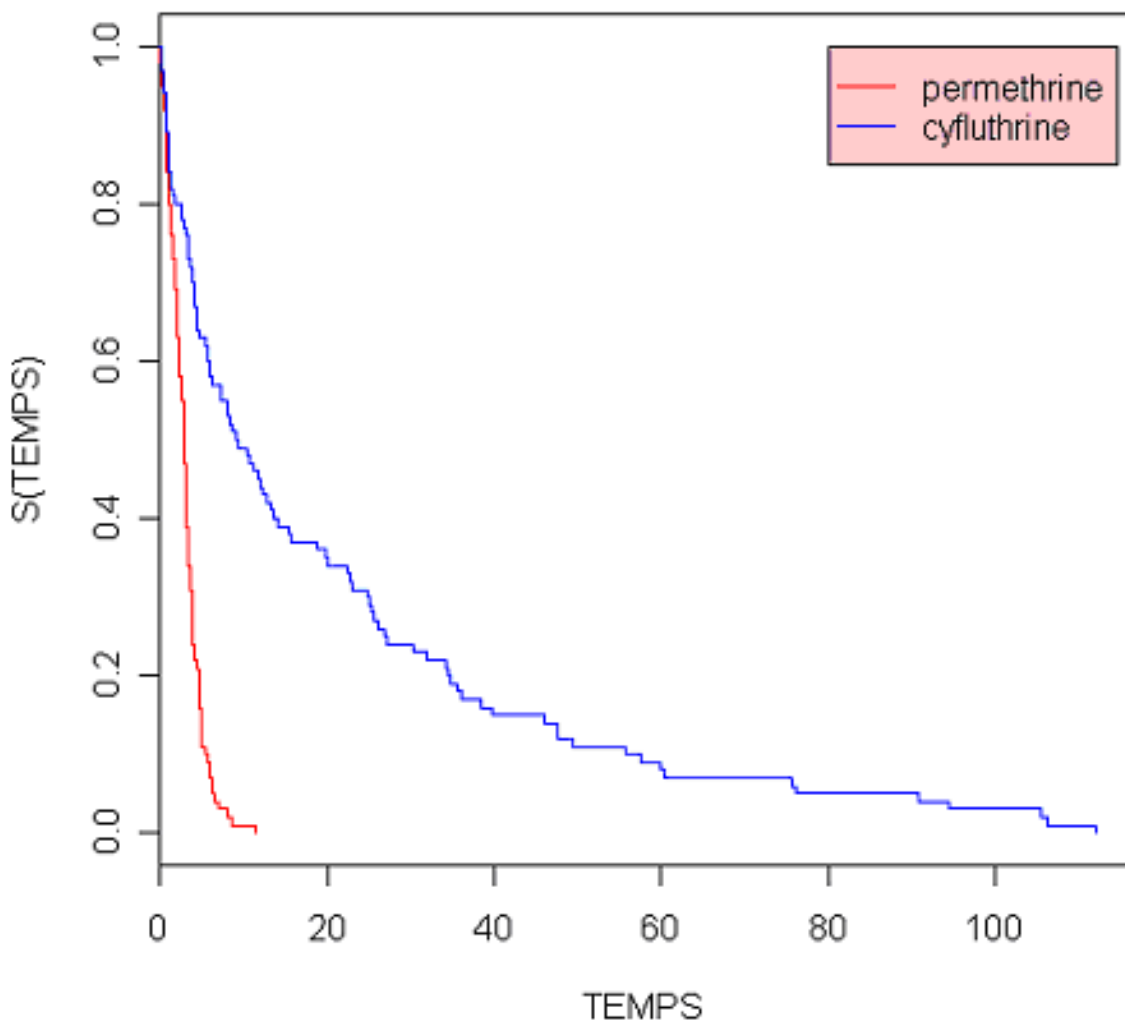


FIG. 3.13 – fig12

Call: survdiff(formula = obj1 inect, rho = 0)  
 N Observed Expected  $(O - E)^2/E$   $(O - E)^2/V$   
 inect=1 100 100 51.7 45.0 75.9  
 inect=2 100 100 148.3 15.7 75.9

Chisq= 75.9 on 1 degrees of freedom, p= 0

La p-value de notre test est nulle, donc on rejette l'hypothèse d'égalité entre les courbes de survie. Ainsi, la perméthrine est plus irritante que la cyfluthrine lors du quatrième mois. C-OPPOSITION DE LA CYFLUTHRINE ET DE LA DELTAMETHRINE EN SOLUTION  
 1- COURBES DE KAPLAN-MEIER 2- TEST DU LOG-RANK

**OPPOSITION CYFLUTHRINE-DELTAMETHRINE\_TABLETT**

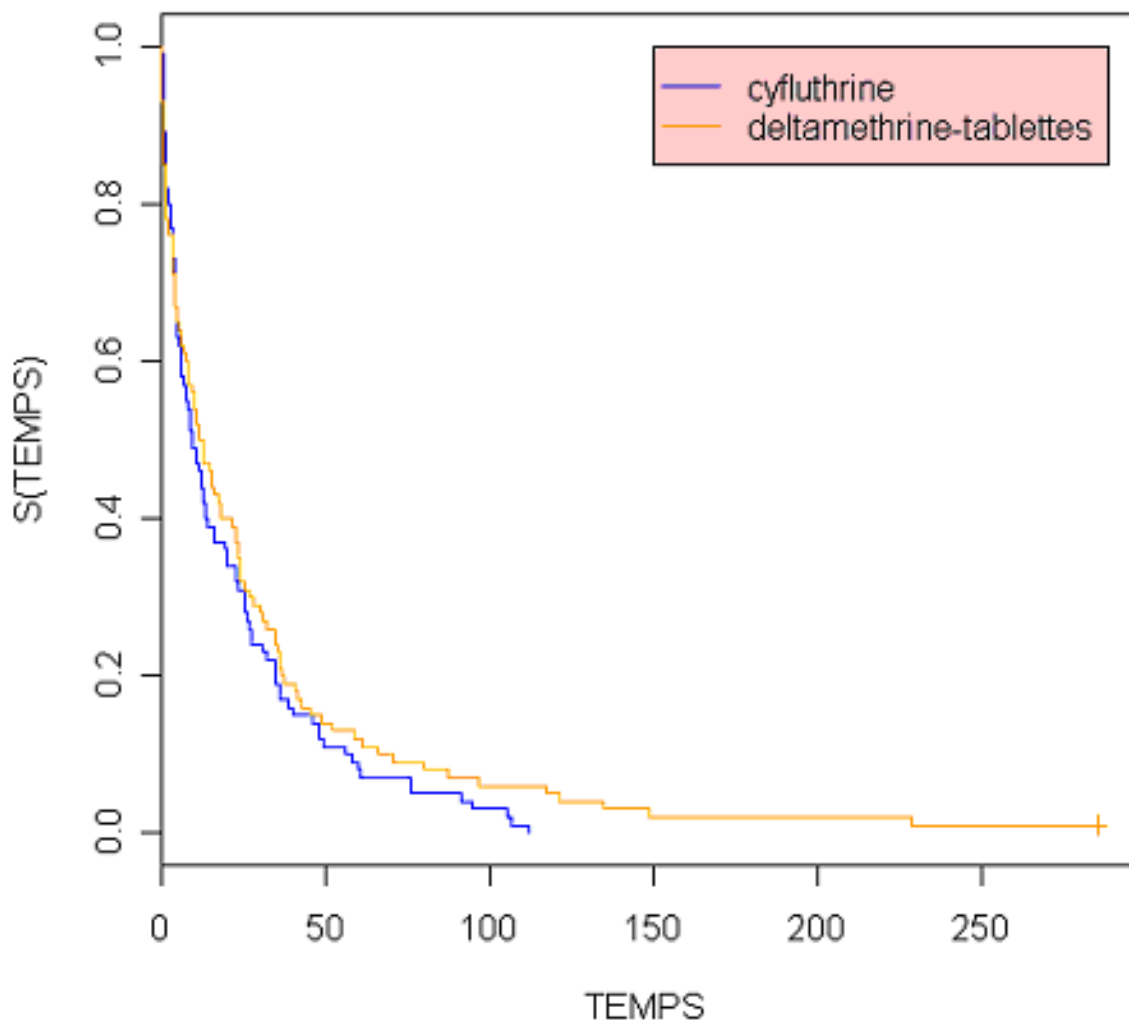


FIG. 3.14 – fig13

```
Call: survdiff(formula = obj1  insect, rho = 0)
      N Observed Expected (O - E)^2/E(O - E)^2/V
insect=1 100 100 92.1 0.672 1.29
insect=2 100 99 106.9 0.579 1.29
```

Chisq= 1.3 on 1 degrees of freedom, p= 0.256

Notre test nous fournit une p-value de  $0.256 > 0.05$ ; on peut ainsi accepter l'hypothèse nulle qui nous dit que les courbes de survie sont égales; donc, lors du quatrième mois, la cyfluthrine et la deltaméthrine en solution ont le même effet d'irritabilité.

D-CONFRONTATION DES DEUX DELTAMETHRINES

1- COURBES DE KAPLAN-MEIER

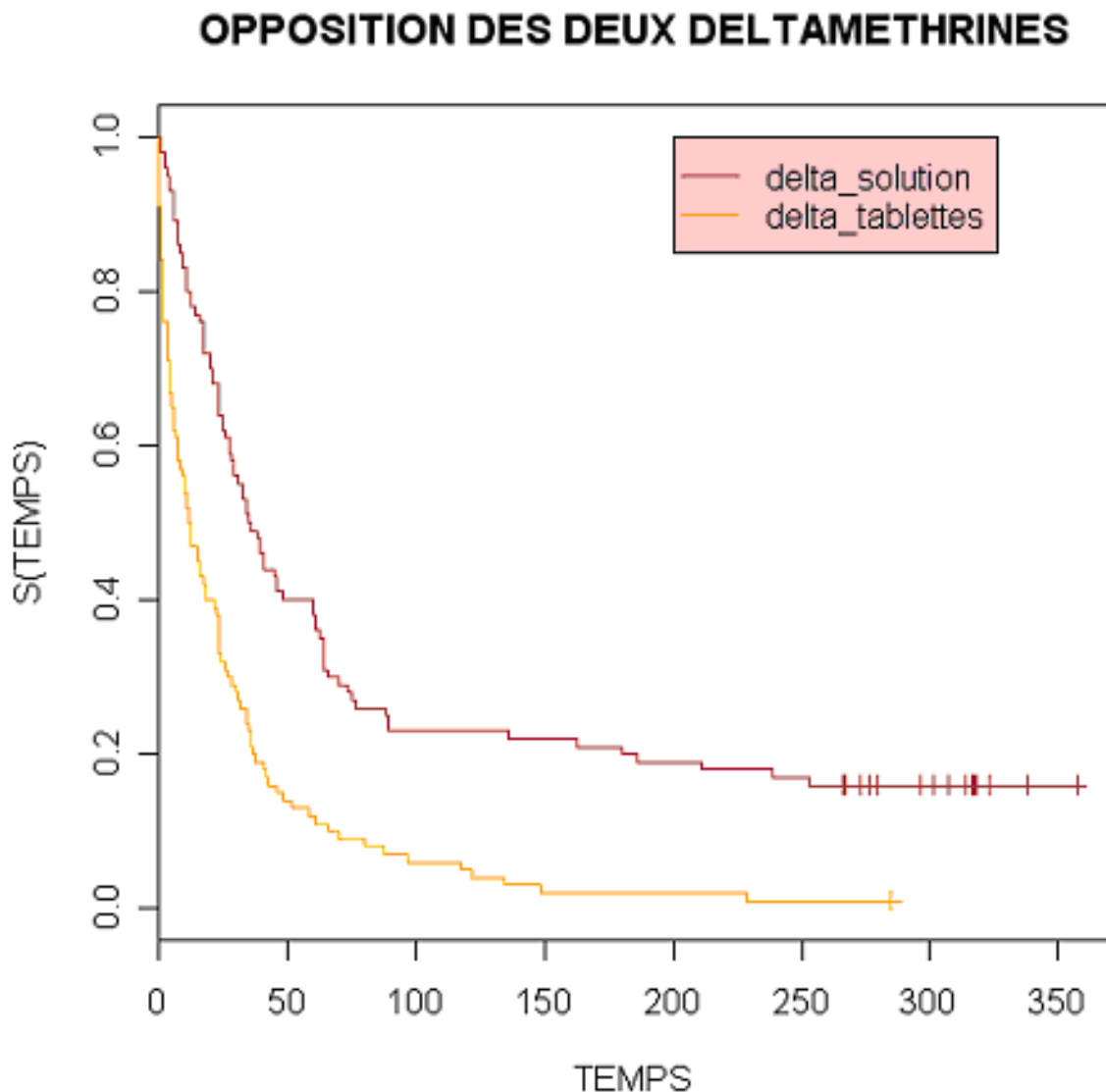


FIG. 3.15 – *fig14*

2-TEST DU LOG-RANK

Call: survdiff(formula = obj1 insect, rho = 0)

N Observed Expected  $(O - E)^2/E$   $(O - E)^2/V$

insect=1 100 84 120 10.8 32.9

insect=2 100 99 63 20.6 32.9

Chisq= 32.9 on 1 degrees of freedom, p= 9.68e-09

Le test du log-rank ici nous fournit une p-value qui est de 9.68e-09 ; on peut ainsi conclure que l'hypothèse d'égalité des courbes de survie est rejetée au profit de l'alternative. Ce qui nous permet de conclure qu'au quatrième mois, la deltaméthrine en tablettes est plus irritante que la deltaméthrine en solution.

---

## CONCLUSION GENERALE

---

Au regard de tout ce qui précède, nous pouvons dire au terme de notre étude que la perméthrine est l'insecticide le plus irritant des quatre; puisque pour chacun des quatre mois elle est plus irritante que les autres insecticides; de même, la cyfluthrine, qui a le même effet d'irritabilité que la deltaméthrine en solution le premier mois est l'insecticide le plus irritant parmi les trois insecticides qui restent; nous constatons aussi que sur les quatre mois, la deltaméthrine en tablette est meilleure que la deltaméthrine en solution seulement au quatrième mois, et n'a le même effet d'irritabilité que lors du premier mois; ceci nous fait dire que la deltaméthrine en solution est meilleure que la deltaméthrine en tablettes sur une période donnée. Les résultats fournis par cette analyse de survie posent tout de même une question dont l'intérêt mérite d'être souligné: puisque la subdivision qui nous permet d'obtenir les temps d'envol des moustiques par intervalles de temps peuvent être découpés de plusieurs manières, les estimations des paramètres qui approchent le plus possible les fonctions de répartition empirique et théorique changent et par conséquent, les résultats numériques changent; il revient donc de se demander quelle est la meilleure subdivision? aussi, l'approche utilisée ici pour modéliser nos phénomènes et résoudre notre problème a été non paramétrique; il convient alors de se pencher vers une approche paramétrique pour une comparaison des méthodes.

---

# ANNEXE

---

Montrons que la matrice:

$(\frac{\partial^2 \ln(f(x, \theta))}{\partial \theta_i \partial \theta_j})_{\hat{\theta}(x)}$  est définie sémi-négative.

$$\ln(f(x, \theta)) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2;$$

ainsi :

$$\frac{\partial \ln(f(x, \theta))}{\partial m} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - m)$$

d'où :

$$\frac{\partial^2 \ln(f(x, \theta))}{\partial^2 m} = -\frac{n}{\sigma^2}$$

et :

$$\frac{\partial^2 \ln(f(x, \theta))}{\partial \sigma^2 \partial m} = -\frac{1}{\sigma^4} \sum_{i=1}^n (x_i - m)$$

on a aussi :

$$\frac{\partial \ln(f(x, \theta))}{\partial \sigma^2} = \frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - m)^2$$

ce qui implique que :

$$\frac{\partial^2 \ln(f(x, \theta))}{\partial^2 \sigma^2} = -\frac{n}{2\sigma^4} - \frac{1}{2\sigma^6} \sum_{i=1}^n (x_i - m)^2$$

en prenant au point  $\hat{\theta}(x)$ , on a :

$$nm = \sum_{i=1}^n x_i, \text{ i.e}$$

$$\sum_{i=1}^n (x_i - m) = 0$$

On obtient donc :

Code R qui estime les deux parametres d'une loi Gamma par la methode du maximum de vraisemblance à partir d'un echantillon observé de cette loi

```
EMV.gamma=function (vec){c1 = sum(vec)c2 = sum(log(vec)) n=length(vec)
```

```
Fonction (G) qui calcule la log-vraisemblance à minimiser
```

```
G=function(p){a = p[1]; lambda = p[2] (n * (a * log(lambda) + lgamma(a))) + ((1 - a) * c2) + (c1/lambda)}
```

```
Estimation initiale par la methode des moments
```

```
m=mean(vec); s2 = var(vec)
```

$a_0 = m^2/s^2$  estimation de  $a$  par la methode des moments

$lambda_0 = s^2/m$  estimation de  $lambda$  par la methode des moments

minimisation effective de la fonction G pour obtenir les EMV recherchés

```
par.gamma=nlm(G,p = c(a0,lambda0))$estimate
```

```
par.gamma = rbind(c(a0,lambda0),par.gamma)
```

Identification des différents estimateurs:

```
colnames(par.gamma) = c("a", "lambda")
```

```
rownames(par.gamma) = c("EMM", "EMV") par.gamma }
```

Code R qui estime les deux parametres d'une loi de weibull par la methode du maximum de vraisemblance à partir d'un echantillon observé de cette loi

```
EMV.WEIBULL=function(vec){
```

fonction (W) qui va calculer la log-vraisemblance à minimiser

```
W=function(parametres){alpha = parametres[1]; lambda=parametres[2]
```

```
-(sum(dweibull(vec,alpha,lambda,log = T))) }
```

Estimation initiale par un algorithme d'initialisation aléatoire

```
initial=function(x){ x/(1 - x) } paraminit=initial(runif(2)) param=paraminit
```

```
param[1] = paraminit[1]; param[2] = paraminit[2]
```

Minimisation effective de la fonction W pour obtenir les EMV recherches

```
par.weibull=nlm(W,p = c(param[1],param[2]))$estimate
```

```
par.weibull=rbind(c(param[1],param[2]),par.weibull)
```

identification des estimateurs

```
colnames(par.weibull) = c("parametre1", "parametre2")
```

```
rownames(par.weibull) = c("INITIAL", "EMV")
```

```
par.weibull
```

```
}
```

Code R qui estime les paramètres qui minimisent l'écart entre la fonction de répartition théorique cherchée et la fonction de répartition empirique.

```
papi=function(N)
```

```
{
```

```
tab=read.table('perm.txt',header = T)
```

```
n=c(3,15,32,31,11,7,0,0,1) x=c(0,1,2,4,8,16,32,64,128,256)
```

```
euclidist=function(theta)
```

```
{ if((theta[1] <= 0)|| (theta[2] <= 0)){g=20 } else{ h=cumsum(n)/(sum(n)) F=pweibull(x,shape = theta[1],scale = theta[2]) g=sqrt(sum((F - h)^2)) } return(g) }
```

```
initial=function(y) { y/(1 - y) }
vap=initial(runif(N * 2)) dim(vap) = c(N,2)
minimisation=function(init) { rs=nlm(euclidist,init,hessian = TRUE) grad=sqrt(sum((rs$g
prop=min((eigen(rs$hessian))$values) if((grad >= .0001 * rs$minimum)|| (prop <= 0)) {
rs$minimum=1000
} return(rs) }
result=apply(vap,1,minimisation) vect=c()
for(iin1 : N) { vect[i] = (result[i])$minimum } minimum=result[vect[i] == min(vect)]

return(minimum)
return(result)
}
```