Biostatistique : Analyse de Survie

A. Latouche

Données Censurées

- Régression logistique : Variable réponse binaire
  - Statut vital : codé en 0 ou 1
  - Présence d'une pathologie
- Si 2 patients décèdent : un à 1 mois et l'autre à 12 mois
- Ils auront le même Statut vital pour une régression logistique

$\Rightarrow$ Perte d'information majeure

Délais d'événement

Ojectifs

- Estimer les fonctions de survie de ses délais
- Tester l'égalité de fonctions de survie entre 2 groupes de patients
- Modèles de régression pour étudier l'effet de facteurs sur le délai de survie

- Délai : **durée** entre survenue de 2 événements (identique ou pas)

- Terminologie *End-Point*, Critère de jugement censuré

Exemples de délais jusqu'à un événement

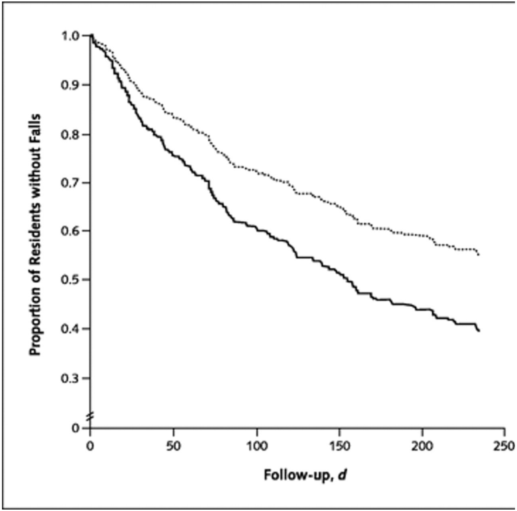
- Première chute

- Apparition de la démence

- Première naissance

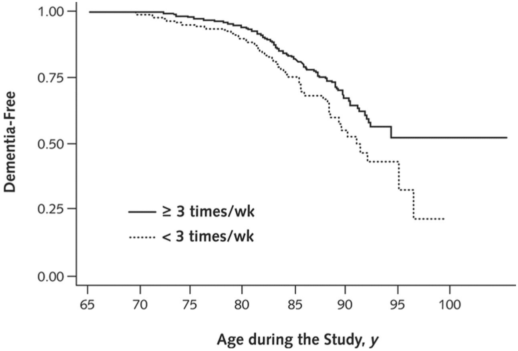
Modèle à 2 états : Origine et Survenu de l'événement d'intérêt

Time to fall

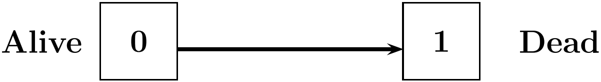


http://www.ncbi.nlm.nih.gov/pubmed/12020141

Time to dementia



http://www.ncbi.nlm.nih.gov/pubmed/16418406



- Que représente la flèche $\rightarrow$?

- Quelle est l'échelle de temps à utiliser (origine) ?
  - Délai depuis l'entrée dans un étude/ délai depuis la naissance
  - Fiabilité d'un voiture : Age ou Kilométrage ?

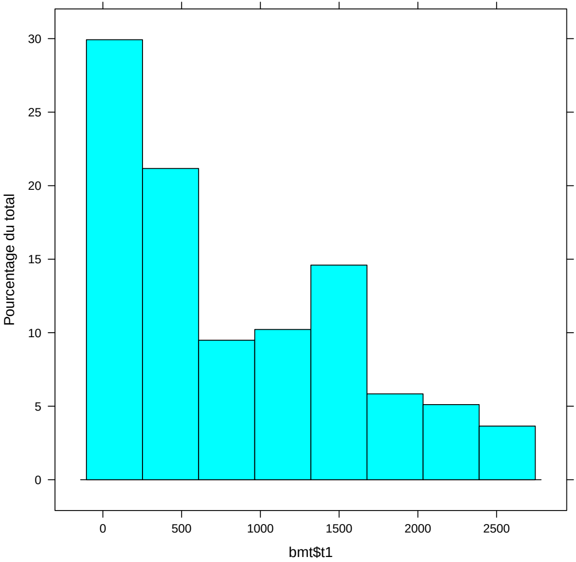
Formalisme : Données censurées à droite

- Délai : modélisé par une variable aléatoire **positive** T
- Pas toujours observé sur la durée de l'étude : censure C

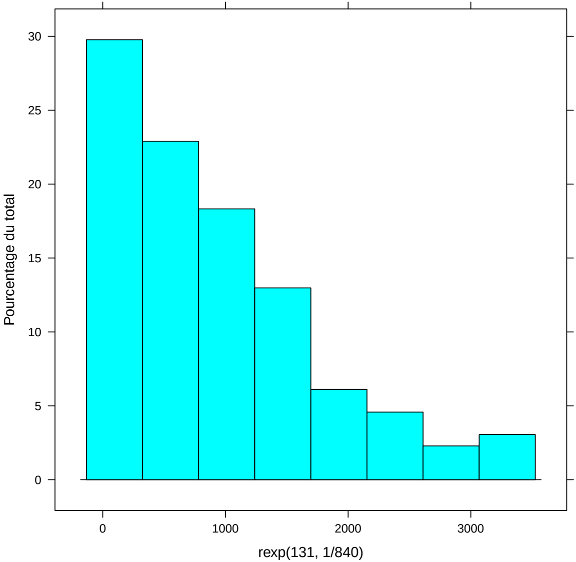
Pour chaque patient on observe un délai et un statut (0 si pas d'événement, 1 si événement)

- Si $T < C$: on observe l'élément et le statut est 1 (l'evt n'est pas censuré)
- Si $T > C$: on n'observe pas l'événement et le statut est 0 (evt censuré)

Données de greffes de moëlle : Délai jusqu'au décès



Données simulées



Modélisation paramétrique pour T

- Loi exponentielle
- Loi de Weibull
- ...

Ces lois continues sont caractérisées par leur densité de probabilité
Rappel pour une v.a. continue

$$\int_{\mathcal{D}_f} f(t)dt = 1$$

et

$$f(t) = \frac{P(t \leq T < t + dt)}{dt}$$

Fonction de répartition et de Survie

Soit T une v.a. positive

- $F(t) = \int_0^t f(u)du = P(T \leq t)$
- F représente la fraction des individus ayant présenté l'événement d'intérêt avant l'instant t

- $S(t) = \int_t^\infty f(u)du = P(T > t)$
- S représente la fraction des individus sans l'événement à l'instant t

$$S(t) = 1 - F(t)$$

Risque instantané (hazard rate)

Le risque instantané

$$\lambda(t) \approx \frac{P(t \leq T < t + dt | T > t))}{dt}$$

où dt est un petit intervalle de temps
d'où

$$\lambda(t) = \frac{P(t \leq T < t + dt)}{dt \ P(T > t))}$$

soit

$$\lambda(t) = f(t)/S(t)$$

Pour décrire T on peut utiliser soit λ soit S

Risque cumulé et Survie

$$\Lambda(t) = \int_0^t \lambda(u)du$$

Comme

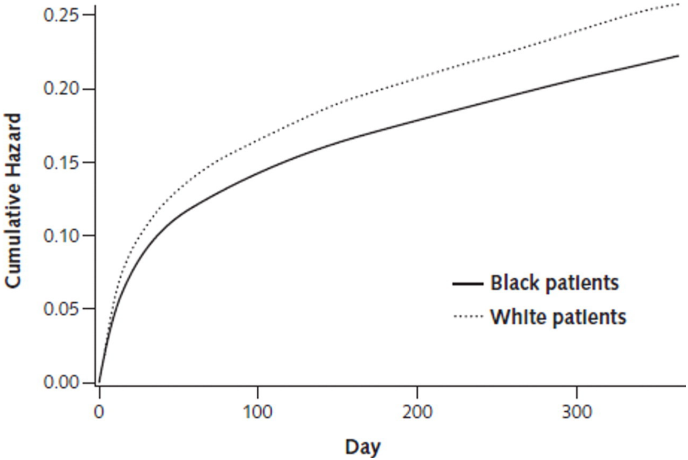
$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{d(log(F(t))}{dt}$$

$$\lambda(t) = -\frac{d(log(S(t))}{dt}$$

$$S(t) = \exp -\Lambda(t)$$

Graphique du risque cumulé : informe sur la variation du risque instantané

Risque cumulé de décès



http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3285233/figure/F1/

Choix d'un modèle paramétrique

La loi de T est décrite par λ

- Si on choisit un risque constant ($=$ L)
- Que valent f, F et S ?

Choix d'un modèle paramétrique

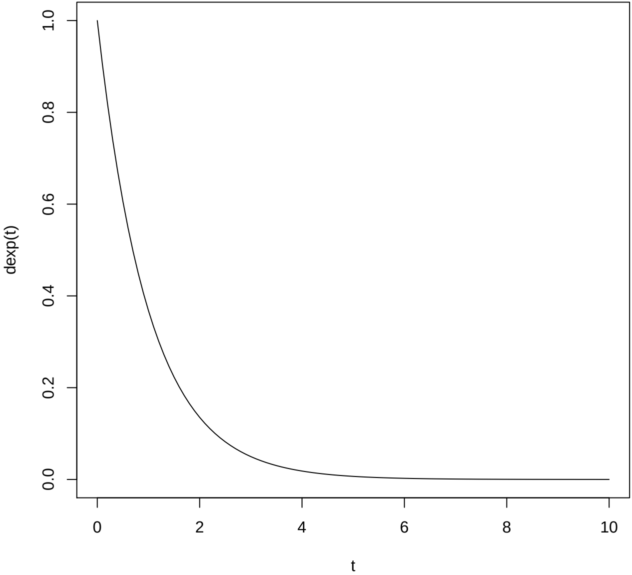
La loi de T est décrite par λ

- Si on choisit un risque constant ($=$ L)
- Que valent f, F et S ?

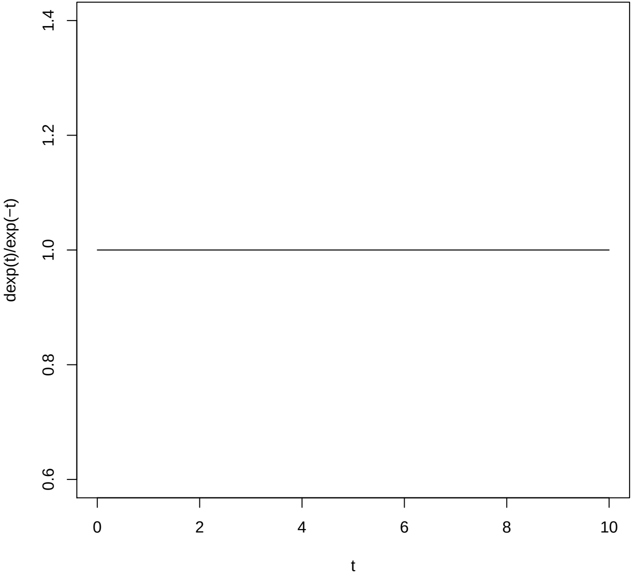
On dit que $T \sim Exp(L)$, Loi exponentielle de paramètre L

- $f(t) = L\exp(-L\ t))$ pout $t \geq 0$
- $S(t) = \exp(-L\ t)$
- $F(t) = 1 - \exp(-L\ t)$

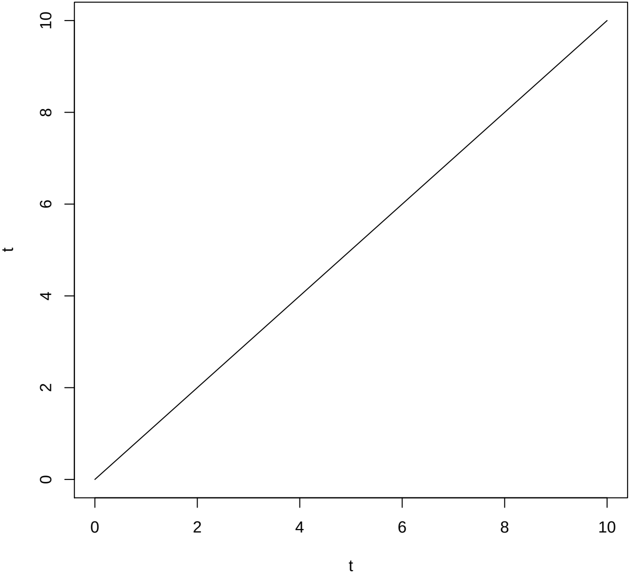
Densité de probabilité Loi Exp(1)



Risque instantané : Loi Exp(1)



Risque Cumulé : : Loi Exp(1)

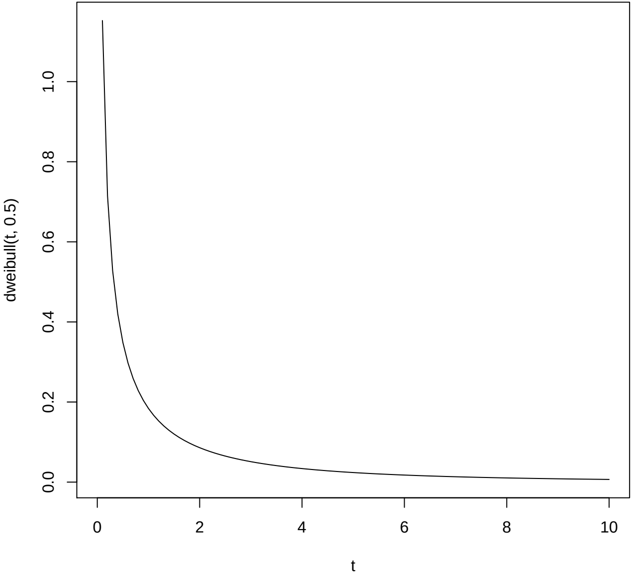


Loi de Weibull

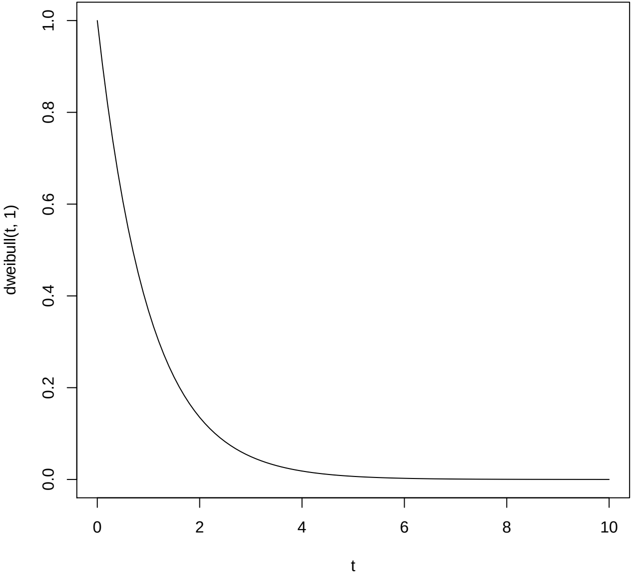
- La fonction de risque est plus *fléxible*
- 2 paramètres *shape* (β) et *scale* (L)
- Risque : $\lambda(t) = L\beta(Lt)^{\beta-1}$
- Survie : $S(t) = exp(-(Lt)^{\beta})$

- Si $\beta = 1$: $Exp(L)$ et risque constant
- Si $\beta < 1$ risque décroissant
- Si $\beta > 1$ risque croissant

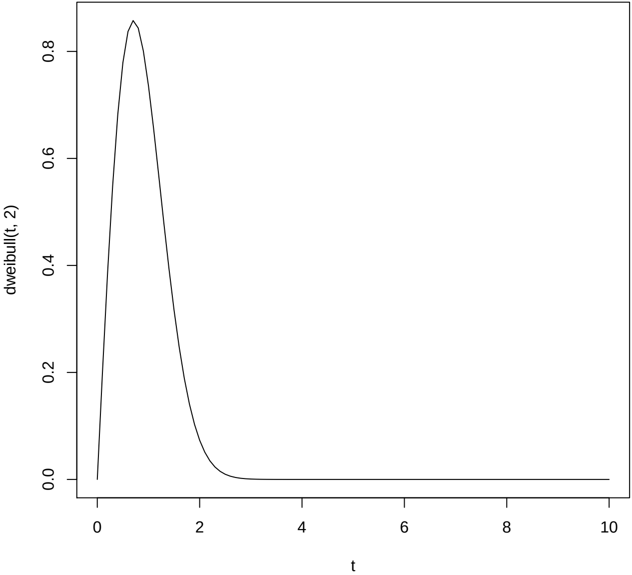
Densité de la loi Weibull(0.5,1)



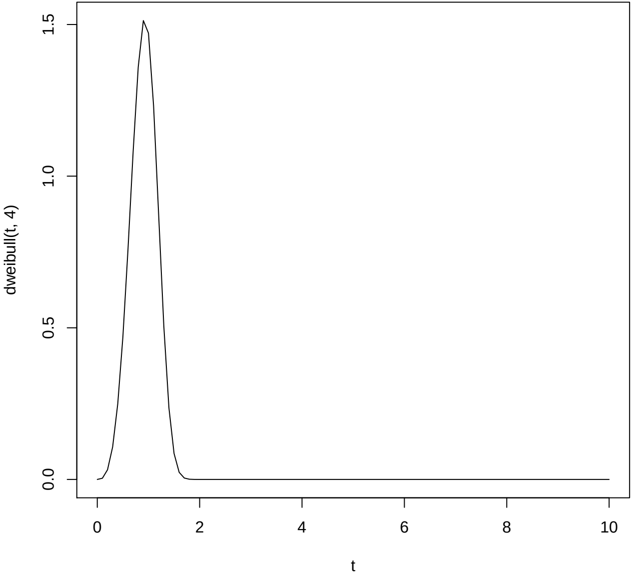
Densité de la loi Weibull(1,1)



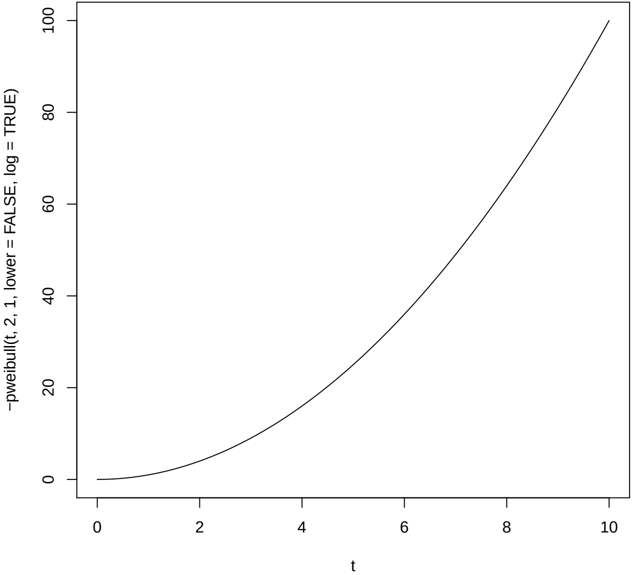
Densité de la loi Weibull(2,1)



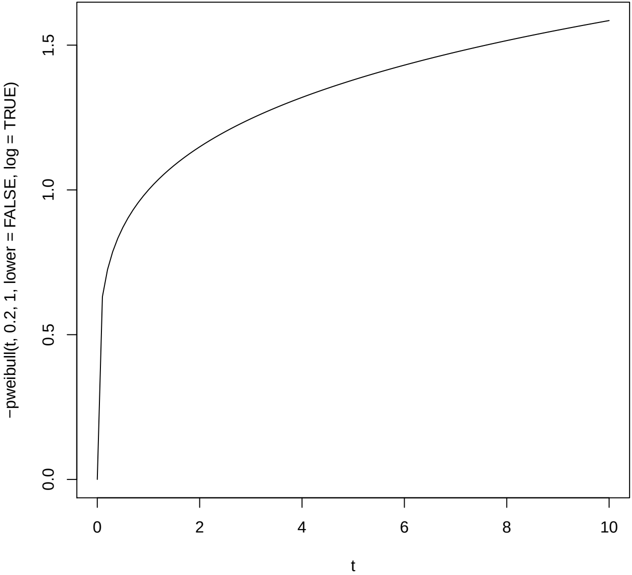
Densité de la loi Weibull(4,1)



Risque Cumulé $\beta > 1$



Risque Cumulé $\beta < 1$



Estimateur de Kaplan-Meier

Objectif

Estimer la fonction de survie quand les données sont censurées à droite

Sans censure on estime la fonction de répartition par

$$\hat{F}(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}(T_i \leq t)$$

La méthode de Kaplan-Meier découle du raisonnement simple suivant :

Ne pas encore avoir subi l'évènement à l'instant t c'est :

- ne pas l'avoir subi juste avant t
- et ne pas le subir en t.

Estimateur de Kaplan-Meier

Soient $t_1 < \ldots < t_k$ les différents temps d'évènements observés, alors on peut définir :

- d_j : nombre de sujets subissant l'évènement au temps t_j
- c_j : nombre de sujets *censurés* dans l'intervalle $[t_j, t_{j+1}[$
- n_j: nombre de sujets à risque au temps t_j (effectif à risque)

Calcul des individus à risque

$$n_j = n_{j-1} - c_{j-1} - d_{j-1}$$

Estimateur de Kaplan-Meier

ne pas encore avoir subi l'évènement en t_j c'est

- ne pas encore avoir subi l'évènement en t_{j-1}
- et ne pas le subir en t_j,

En utilisant les probabilités conditionnelles on obtient

$$\begin{aligned}
\hat{S}(t_j) &= \hat{P}(X > t_j) \\
&= \hat{P}(X > t_j | X > t_{j-1}) \times \hat{P}(X > t_{j-1}) \\
&= \hat{P}(X > t_j | X > t_{j-1}) \times \hat{P}(X > t_{j-1} | X > t_{j-2}) \\
& \quad \times \hat{P}(X > t_{j-2}) \\
& \quad \ldots \hat{P}(X > t_2 | X > t_1)
\end{aligned}$$

Chaque probabilité $P(X > t_j | X > t_{j-1})$ est estimée par $(n_j - d_j)/n_j$

$$\hat{S}(t) = \prod_{t_j < t} \frac{n_j - d_j}{n_j}$$

Présentation des données censurées (kidney package KMsurv)

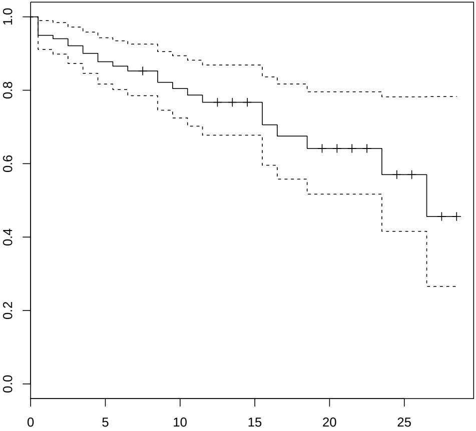
time : delay d'infection en mois

type : intervention pour la pause du catheter

	time	delta	type
1	1.50	1	1
2	3.50	1	1
3	4.50	1	1
4	4.50	1	1
5	5.50	1	1
6	8.50	1	1

table(kidney$delta) nous fournit les effectifs en fonction du status
soit 93 patients avec status=0 et 26 patients avec status=1
Soit 78% de données censurées

Estimation de la fonction de survie



package : KMsurv, data(kidney)

Données Censurées : Estimation de la fonction de survie $S(t) = P(T > t)$

Sujet	Delai	Status
1	2	1
2	3	1
3	3	1
4	4	1
5	6	1

Compléter le tableau suivant

Données Censurées : Estimation de la fonction de survie $S(t) = P(T > t)$

Sujet	Delai	Status
1	2	1
2	3	1
3	3	1
4	4	1
5	6	1

Compléter le tableau suivant

Période	Nb de décès	Individus à risque	Proba	Taux

Données Censurées : Estimation de la fonction de survie $S(t) = P(T > t)$

Sujet	Delai	Status
1	2	1
2	3	1
3	3	1
4	4	1
5	6	1

Compléter le tableau suivant

Période	Nb de décès	Individus à risque	Proba	Taux
$[0; 2[$	0	5	$0/5$	1

Données Censurées : Estimation de la fonction de survie $S(t) = P(T > t)$

Sujet	Delai	Status
1	2	1
2	3	1
3	3	1
4	4	1
5	6	1

Compléter le tableau suivant

Période	Nb de décès	Individus à risque	Proba	Taux
$[0;2[$	0	5	0/5	1
$[2;3[$	1	5	1/5	1-1/5=0.8

Données Censurées : Estimation de la fonction de survie $S(t) = P(T > t)$

Sujet	Delai	Status
1	2	1
2	3	1
3	3	1
4	4	1
5	6	1

Compléter le tableau suivant

Période	Nb de décès	Individus à risque	Proba	Taux
$[0; 2[$	0	5	0/5	1
$[2; 3[$	1	5	1/5	1-1/5=0.8
$[3; 4[$	2	4	2/4	(1-1/5)(1-1/2)

Données Censurées : Estimation de la fonction de survie $S(t) = P(T > t)$

Sujet	Delai	Status
1	2	1
2	3	1
3	3	1
4	4	1
5	6	1

Compléter le tableau suivant

Période	Nb de décès	Individus à risque	Proba	Taux
$[0; 2[$	0	5	0/5	1
$[2; 3[$	1	5	1/5	1-1/5=0.8
$[3; 4[$	2	4	2/4	(1-1/5)(1-1/2)
$[4; 6[$	1	2	1/2	0.4*1/2

Données Censurées : Estimation de la fonction de survie $S(t) = P(T > t)$

Sujet	Delai	Status
1	2	1
2	3	1
3	3	1
4	4	1
5	6	1

Compléter le tableau suivant

Période	Nb de décès	Individus à risque	Proba	Taux
$[0; 2[$	0	5	0/5	1
$[2; 3[$	1	5	1/5	1-1/5=0.8
$[3; 4[$	2	4	2/4	(1-1/5)(1-1/2)
$[4; 6[$	1	2	1/2	0.4*1/2
$[6;\ [$	1	1	1/1	0

Données Censurées : Estimation de la fonction de survie $S(t) = P(T > t)$

Sujet	Delai	Status
1	2	1
2	3	1
3	3	1
4	4	1
5	6	1

Compléter le tableau suivant

Période	Nb de décès	Individus à risque	Proba	Taux
$[0; 2[$	0	5	0/5	1
$[2; 3[$	1	5	1/5	1-1/5=0.8
$[3; 4[$	2	4	2/4	(1-1/5)(1-1/2)
$[4; 6[$	1	2	1/2	0.4*1/2
$[6;\ [$	1	1	1/1	0

Exercice

La médiane de survie est de 12 mois chez un groupe de souris. On suppose que l'on peut modéliser le délai de survie par une loi exponentielle.

- Quelle est la probabilité qu'une souris encore en vie au bout d'un an, décède dans la 2eme année ?

- $P(X < 24|X > 12) - = P(12 < X < 24)/P(X > 12) = (P(X < 24) - P(X < 12))/0.5$

- Propriété la loi exponetielle est sans mémoire. En effet $P(X > b|X > a) = P(X > b)/P(X > a)$ or $P(X > a) = exb(-\lambda a)$ donc $P(X > b)/P(X > a) = exp(-\lambda b)/exp(-\lambda a)$

$$P(X > b|X > a) = P(X > b - a)$$