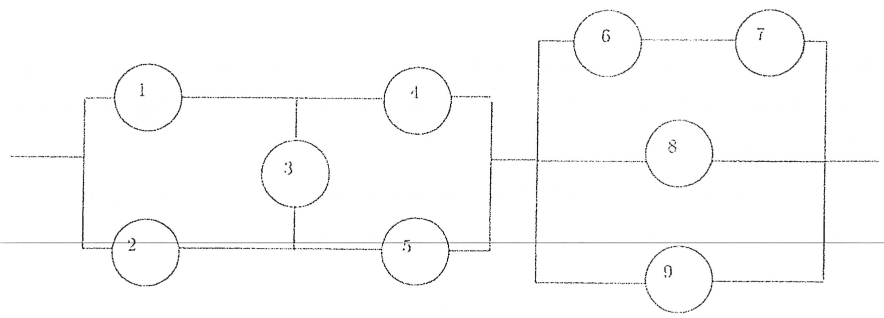
TP1

On considère le système complexe suivant composé de deux blocks en série :



1. Déterminer la fonction de structure des différents blocs du système, ainsi que celle du système ϕ. Ecrire une fonction prenant en argument les états (binaires) x_i des composants et renvoyant l'état $\phi(x_1, ..., x_9)$ du système.
2. Ecrire une fonction prenant en argument les fonctions de survie des composants et renvoyant la fonction de survie $\bar{F}$ de durée de vie T du système.
3. Tracer cette fonction de répartition lorsque les durées de vie suivent des lois de Weibull dont on fera varier les paramètres λ et β.
4. Prendre des coefficients β des lois de Weibull supérieurs à 1 strictement, et tracer le graphe de $t \to -\ln \bar{F}(t)/t$. Que peut on dire de la courbe ? Tracer quelques courbes pour des valeurs de β inférieures à 1 et comparer.
5. Comment peut on simuler une variable aléatoire de fonction de répartition G à partir de la loi uniforme, si G^{-1} est exprimable simplement ? Utiliser cette méthode pour écrire une fonction renvoyant une réalisation de $\phi(X_t)$ sur un intervalle $[0, a]$ (a pris en argument, de même que les paramètres des lois de Weibull éventuels). Tracer quelques trajectoires de $t \mapsto \phi(X_t)$.

6. Ecrire une fonction donnant une estimation de la durée de vie T du système. On pourra utiliser le fait que l'on a

$$T \approx t_N, \quad N = \max\{n| \ \phi(X_{t_i}) = 1\},$$

où $(t_i)_{i=1,\dots,n}$ est une subdivision de pas très petit et n est grand. Ecrire de même deux fonctions donnant une estimation des durées de vie T_1 et T_2 des blocks 1 et 2.

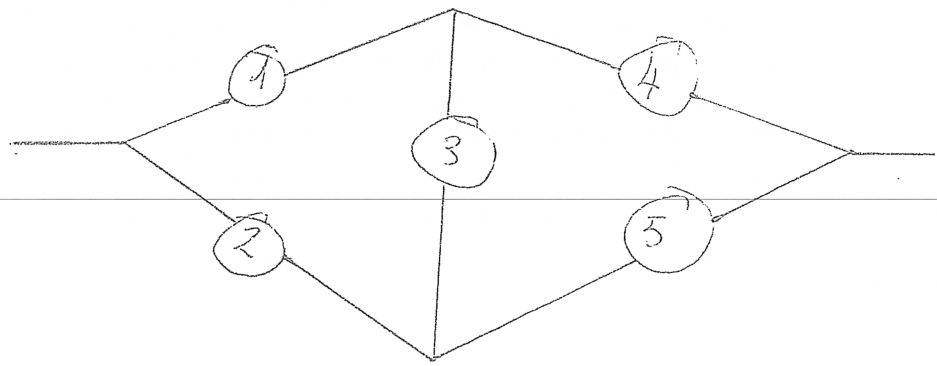
7. Estimer alors $\mu = E(T)$ à l'aide de la loi des grands nombres. Vérifier expérimentalement que lorsque les paramètres β sont supérieurs à 1, on a

$$\mu \geq \left(\frac{1}{\mu_1} + \frac{1}{\mu_2}\right)^{-1} \tag{1}$$

où μ_1 et μ_2 sont les durées de vies moyennes des blocks 1 et 2. Vérifier que l'inégalité (1) n'est pas toujours vérifiée lorsque les paramètres sont inférieurs à 1.

TP2

Exercice 1 : On considère le système complexe suivant :



On suppose cette fois ci que les temps successifs de vie et de réparation sont indépendants et suivent des lois exponentielles de paramètres λ et μ, pour ces composants.

1. Ecrire une fonction prenant en argument un vecteur d'états du système et retournant la valeur de l'état du système.
2. On se propose de simuler l'état du système sur un intervalle de temps.
   (a) Ecrire une fonction prenant en argument λ, μ ainsi que le temps de simulation T, et renvoyant une matrice à six lignes dont la première ligne est la succession des instants de panne et de réparation du premier composant jusqu'au premier instant de réparation supérieur à T, et dont la deuxième ligne vaut -1 ou $+1$ suivant que l'instant en question est un instant de panne ou de réparation, les autres lignes étant remplies de zéros.
   (b) Ecrire (en faisant beaucoup de copier-coller !) quatre autres fonctions renvoyant des matrices de six lignes et dont les lignes numéro 3,..., 6 décrivent le comportement de panne ou réparation de chacun des quatre autres composants, la première ligne comprenant les instants successifs de panne et réparation.
   (c) Ecrire une fonction prenant en argument λ, μ ainsi que le temps de simulation T, et renvoyant une matrice à six lignes composée des cinq matrices obtenues grâce aux fonctions précédentes concaténées

et triées selon la première ligne (celle des instants de panne et réparation). Penser à utiliser `gsort` avec l'option `'lc'`.

(d) Ecrire une fonction renvoyant les instants de panne et remise en fonctionnement du système. Plus précisément, elle prendra en argument λ, μ ainsi que le temps de simulation T, et renverra le vecteur des instants de changements d'états, les vecteurs d'états correspondants et le vecteur d'état du système.

3. Ecrire une fonction prenant en argument les même paramètres, et renvoyant le nombre de pannes N_T sur l'intervalle $[0, T]$. Déterminer $\lim_{T\to\infty} N_T/T$.

Exercice 2 : On considère à présent le système complexe du TP précédent. Les temps de vie des composants suivent à présent des lois de Weibull de paramètres λ et β. On se place dans le cadre où le système est remis à neuf ou réparé, avec inspection éventuelle au bout de T unités de temps si une panne n'a pas eu lieu ou si elle n'a pas été détectée. Soit ξ_i la durée entre la $i-1$ème et $i$ème remise à neuf du système, et ξ_i^* la $i$ème durée d'occupation.

1. Déterminer l'expression de la disponibilité $A(T) = E(\xi_i^*)/E(\xi_i)$ du système en fonction de T.
2. En prenant quelques valeurs pour λ, β, t_s, t_e, t_p, α, et en utilisant la fonction de répartition $F(t)$ déterminée dans le TP no 1, tracer la courbe de la disponibilité du système $T \mapsto A(T)$ et déterminer la valeur optimale T^* de T maximisant la disponibilité asymptotique du système. On utilisera pour les intégrales $\int_0^T \bar{F}(s)ds$ une approximation du type $\frac{1}{N}\sum_{i=1}^{N} \bar{F}(x_i)$ avec $0 < x_1 < \ldots < x_N = T$ et $x_i - x_{i-1} = 1/N$. On pourra également tracer une nappe en faisant également varier α en utilisant `plot3d`.

Projet 1

1. En vous inspirant de vos recherches sur internet, trouver (sans y passer vos journées...) un système complexe ayant une dizaine d'éléments. Tracer son diagramme et déterminer sa fonction de structure.
2. En prenant par exemple des temps de durée de vie indépendants suivant des lois de Weibull, écrire une fonction donnant une trajectoire de l'état du système entre 0 et T, ainsi qu'une fonction donnant une estimation de la fonction de survie du systéme.
3. Dans le cas où les composants sont IFRA, vérifier et illustrer le fait que le système est IFRA.
4. On se place dans le cas où des réparations sont possibles pour chaque composant. Les durées de vie (T_i) et de réparations (R_i) sont i.i.d., indépendantes et de lois de Weibull (de paramètres différents). Ecrire une fonction qui donne l'état du système sur un intervalle $[0, w]$ et afficher quelques simulations.
5. Ecrire une fonction qui donne le nombre de pannes $N[0, w]$ sur un intervalle $[0, w]$. De même, écrire une fonction qui donne le temps de fonctionnement Y_w sur cet intervalle. Vérifier si des lois des grands nombres et un TCL du type

$$\frac{N[0,w]}{w} \xrightarrow{p.s.} m, \quad \frac{Y_w}{w} \xrightarrow{p.s.} l, \quad \frac{N[0,w]-mw}{\sqrt{w}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma^2)$$

   lorsque $w \to \infty$, sont plausibles (cela dépendra peut être du système complexe).
6. On suppose que chaque composant est soumis à maintenance, plutôt que réparations. Autrement dit, on décide que le système est remis en état au bout d'un temps fixe T, et que le temps de remise à neuf dure δ, pendant lequel le système n'est pas opérationnel. Après la remise à neuf tous les composants sont opérationnels. Comparer le temps de fonctionnement moyen sur un intervalle fixé grand, en prenant des paramètres tels que les deux solutions semblent être équivalentes.

Master2 Stat
Université de Franche Comté

Fiabilté

Projet R

1. Eventuellement à l'aide d'une recherche sur internet, donner le diagramme d'un système complexe ayant une dizaine de composants au minimum, et donner sa fonction de structure ϕ.
2. En prenant des durées de vie de lois classiques facilement simulables (ex. Weibull), effectuer quelques simulations et tracer quelques trajectoires de l'état du système. Estimer le temps de vie moyen $E(T)$, avec intervalle de confiance.
3. On se concentre sur un composant en particulier, que vous choisirez comme bon vous semble, en essayant d'en prendre un qui a un rôle "critique". On fait intervenir un technicien qui remet le composant en marche tous les intervalles de temps δ. S'il est en panne il le répare, s'il fonctionne il le laisse en l'état, et dans les deux cas cela induit un coup de déplacement C.
   (a) Simuler et tracer quelques trajectoires de $t \mapsto \phi(X_t)$.
   (b) On considère la fonction de récompense
   $$R(\delta) = aE(T) - CE(N)$$
   où a est un coefficient représentant le gain associé au temps de fonctionnement de la machine, et N le nombre d'intervention du technicien avant l'instant T de panne du système. Estimer numériquement $R(\delta)$ et essayer de maximiser cette quantité par rapport à δ (choisir des valeurs de a, C plausibles en fonction du contexte).

Estimation du temps que vous devrez passer sur ce projet : Q1 : 2 heures maximum, Q2 : 1 heure en reprenant le TP, Q3 : 3 à 5 heures. Compter environ deux heures pour la rédaction du rapport et la préparation de la soutenance, qui durera environ 15 minutes, questions comprises.

Projet 3 & 4

Exercise 1: Familiarization with the function survfit

The goal of this first exercise is to get familiarized with the function `survfit` of the software `R` that provides the Kaplan-Meier and Nelson-Aalen estimations of the survival function and of the cumulative hazard rate function. We work on the example `lung` already in `R`.
Write and comment the following commands:

```
library(survival)
help(lung)
kmfit<-survfit(Surv(time,status)~ 1,data=lung,conf.type="plain",type='kaplan-meier')
print(kmfit)
summary(kmfit)
plot(kmfit)
plot(kmfit,mark.time=F,xscale=365.25,xlab="Time (in years)",ylab="Survival S(t)")
legend(1,0.8, c("Kaplan-Meier function", "95\% pointwise CI"), lty=1:2)
fhfit<-survfit(Surv(time,status)~1,data=lung,conf.type="plain",type='fh')
plot(kmfit,mark.time=F,xscale=365.25,xlab="Time (in years)",ylab="Survival S(t)")
lines(fhfit,lty=3,mark.time=F,xscale=365.25,col="red")
plot(kmfit\$time,kmfit\$surv-fhfit\$surv)
naH =-log(fhfit\$surv)
time= fhfit\$time
plot(time,naH,type="s",ylab="Cumulative risk H(t)",xlab="Time (in months)")
```

Exercise 2

The following data come from a clinical trial led by Freireich, in 1963. The goal was to compare the remission durations (in weeks) of patients that suffer from leukemia. The patients are divided into two subgroups: some of them received a medicine (6-MP) and the others a placebo. The results are presented in the following tabular:

6-MP	6	6	6	6^+	7	9^+	10	10^+	11^+	13	16
	17^+	19^+	20^+	22	23	25^+	32^+	32^+	34^+	35^+	
Placebo	1	1	2	2	3	4	4	5	5	8	8
	8	8	11	11	12	12	15	17	22	23	

The patients with a + sign correspond to lost subjects at the considered time of observation: they are censored, "excluded-alive" of the study and one only knows that their remission duration is greater than the observed delay.

1. Compute the Kaplan-Meier estimator of the survival function S. Estimate its variance.

   One may use the following tabular to lead the calculus:

Time of relapse	Number of relapses	Censoring in $[T_{(i)}, T_{(i-1)}[$	At risk numbers at $T_{(i)}$	Conditional proba-bility	Survival probability without relapse
$T_{(i)}$	m_i	c_{i-1}	n_i	$(n_i - m_i)/n_i$	$\hat{S}_{n,KM}(T_{(i)})$
Placebo					
1					
2					
3					
4					
5					
8					
76					
11					
12					
15					
17					
22					
23					
6-MP					
6					
7					
10					
13					
16					
22					
23					

2. Recover the previous results using `R` and plot the graph of the estimated survival function with respect to the time.

3. In the group that has received the placebo, the remission times are:

$$1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23.$$

   Estimate using `R` the survival functions in each group using the Kaplan-Meier estimator and plot them.

4. Determine the Breslow and the Nelson-Aalen estimation of the cumulative hazard rate function H.

   One may use the previous and following tabulars to lead the calculus:

Time of relapse	Number of relapses	At risk numbers at $T_{(i)}$	Nelson proportion h	Nelson estimation	Kaplan-Meier estimations	
$T_{(i)}$	m_i	n_i	m_i/n_i	$\hat{H}_{n,NA}(T_{(i)})$	$\hat{S}_{n,KM}(T_{(i)})$	$\hat{H}_{n,BR}(T_{(i)})$
Placebo						
1						
2						
3						
4						
5						
8						
76						
11						
12						
15						
17						
22						
23						
6-MP						
6						
7						
10						
13						
16						
22						
23						

Exercise 3

From February 1998 to February 2001, 29 patients that suffered from a severe viral hepatitis were admitted in a therapeutic trial of 16 weeks. The goal was to compare the effect of a therapy with steroids. The patients received randomly the treatment or the placebo. The survival times (in weeks) of the goups of the 14 patients treated are

$$1, 1, 1, 1^+, 4^+, 5, 7, 8, 10, 10^+, 12^+, 16^+, 16^+ 16^+.$$

1. No assumption has been done on the survival time distribution under treatment.
   (a) Estimate cumulative risk H with the Nelson-Aalen estimator.
   (b) Deduce the Harrington and Fleming estimator of S.
   (c) Determine the Kaplan-Meier estimator of S.
   (d) Represent these two estimators of S on a same figure using `R`.
2. Now we assume that the survival time is exponentially distributed with parameter λ.
   (a) Estimate λ by the maximum likelihood method.
   (b) Deduce the estimation of the probability to survive more than 16 weeks.
   (c) Estimate the median of the survival time.
3. Represent these three estimators of S on a same figure using `R`. Comment.

Exercise 4

1. Generate a sample of size 100 of a random variable X exponentially distributed with parameter $\lambda = 1.1$. Represent on the same figure the theoretical and empirical survival functions of X using R.

2. Generate a sample of size 100 of the pair $(T = \min(X, C), \delta)$, where $X \sim \mathcal{E}(1.1)$, $C \sim \mathcal{E}(1)$ and $\delta = \mathbb{1}_{\{X \leqslant C\}}$.

   (a) Compute the Kaplan-Meier survival function estimation obtained considering the whole observations.

   (b) Determine the estimation of the survival function by the maximum likelihood method on the whole observations.

   (c) Represent on the same figure the theoretical survival function of X, its Kaplan-Meier estimation and the MLE estimation.

3. Select the uncensored observations.

   (a) Compute the Kaplan-Meier survival function estimation obtained considering the uncensored sample.

   (b) Estimate the survival function by the maximum likelihood method on the uncensored observations.

   (c) On the previous figure, represent these two functions.

4. Same question by making vary the sample size. Conclusion?

5. CI comparisons

   We work on the whole sample. Represent on the same figure the theoretical survival function of X and its Kaplan-Meier estimation.
   Add the confidence intervals of types "plain", "log" and "log-log" for $S(t)$ on three different figures.
   To which formulas do these intervals correspond?
   Conclusion?