LA FONCTION DE REPARTITION POUR QUOI FAIRE ?

Pascale POMBOURCQ
Irem de Toulouse

A. INTRODUCTION

L'idée de cet article part de discussions quelquefois animées sur l'utilité de la fonction de répartition en classe de Terminale. Ces discussions ont eu lieu, tant au sein du groupe probabilités-statistiques de l'IREM de Toulouse que lors de l'animation de stages sur les probabilités.

Deux points de vue se sont le plus souvent opposés :

— Pour une loi discrète, la fonction de répartition n'apporte rien quant à la compréhension de la loi qui lui est associée, c'est un objet compliqué et comme tout objet compliqué, si on veut le réinvestir par la suite, il faut le faire fonctionner suffisamment, ce qui est rarement le cas par manque de temps. Elle n'est alors qu'un objet d'apprentissage.

— En Terminale, la fonction de répartition est quasiment l'unique occasion de rencontrer des fonctions non continues et plus précisément des fonctions en escalier (il est à noter que la notion de continuité disparaît des nouveaux programmes de Terminale S et ne figurait pas en tant que telle dans les autres Terminales). De plus elle permet de mettre en place des notions qui seront utilisées dans les classes post-bac.

Lors de notre enseignement de la fonction de répartition en classe de Terminale, il nous a semblé que les difficultés rencontrées par les élèves provenaient essentiellement du maniement des fonctions en escalier.

Le chapitre sur les variables aléatoires ne pose en général aucun problème de compréhension aux élèves. Cette notion, ainsi que les notions qui lui sont associées, leur paraissent tout à fait naturelles. Et bien qu'ils sachent calculer la probabilité de l'événement $[X \leq a]$ pour un nombre donné a et que cette écriture leur vienne spontanément (si X est une variable aléatoire désignant un gain, ils traduisent très rapidement «gagner moins de 20 francs» par $[X \leq 20]$), la liaison avec la fonction de répartition est toujours difficile. Dans cette dernière, le *a* devient *x* et se met à varier. C'est en fait toute la difficulté du passage du discret au continu. Ce qui leur pose alors problème est la conjonction de l'intervalle et de la probabilité. En effet : si

$$X(\Omega) = \{ x_1 , x_2 , \dots , x_n \} \text{ où } x_i < x_{i+1} ,$$

pour exprimer la fonction de répartition F, il faut dans un premier temps positionner x par rapport aux xi : si $x_i \leq x < x_{i+1}$, $F(x) = ?$

Les élèves ne comprennent pas que dans le calcul de F(x) apparaissent les valeurs inférieures à x_i alors que x prend des valeurs comprises entre x_i et x_{i+1} :

$$F(x) = P[X = x_1] + P[X = x_2] + \dots + P[X = x_i]$$

La difficulté ne vient donc pas de l'écriture de l'événement $[X \leq x]$ mais du sens donné à x, c'est-à-dire que

$$[X \leq x] = [X = x_1 \text{ ou } X = x_2 \text{ ou } \dots \text{ ou } X = x_i]$$

Il serait peut-être préférable, afin d'aider nos élèves à surmonter cette difficulté, de travailler davantage sur le sens de l'événement, plutôt que sur le calcul de sa probabilité elle-même. C'est ce que nous avons essayé de faire dans la partie I qui est proposée un peu plus loin.

L'attitude des enseignants la plus fréquente face à ce type de difficultés nous a paru être la présentation rapide des définitions et propriétés, et le passage immédiat à quelques tracés de cette fonction où les "trucs" sont vite décelés par nos élèves. Ils ont en effet très vite repéré que la première marche se situe sur l'axe des abscisses, la dernière sur la droite d'équation y = 1 , et que pour obtenir les autres valeurs de F(x), on ajoute les probabilités case par case (à partir du tableau donnant la loi de probabilité). Contrairement à ce que l'on pourrait penser, il ne s'agit pour eux que d'un algorithme dont le sens leur échappe.

Or nous pensons que cela vaut la peine de passer un peu de temps à introduire cette fonction. Nous éviterions ainsi l'appréhension de nos élèves vis-à-vis de cet objet, appréhension qui subsiste souvent dans les classes supérieures quand ils aborderont les variables aléatoires continues.

Un thème, que nous rencontrons dans certaines sections de STS, donne du sens à la fonction de répartition. Il s'agit de la fiabilité : la variable aléatoire étudiée est le temps de bon fonctionnement d'un élément. La fonction de répartition qui lui est associée, devient une fonction de défaillance qui a son utilité propre et dont on utilise concrètement la courbe représentative. De plus la fiabilité est utilisée dans des domaines très variés tels que l'aéronautique, le nucléaire, l'informatique, la médecine.

Cet article se décompose en trois parties, qui sont des propositions d'activités en classe de Terminale.

Partie I : un exemple de présentation de la fonction de répartition conformément au programme des classes où figurent les variables aléatoires.

Partie II : un premier TP qui a pour but, à partir de l'étude d'un échantillon, de tracer la courbe représentative d'une estimation de

la fonction de fiabilité.
— dans un premier temps, point par point
— ensuite à partir de l'étude du nuage de points, retrouver la fonction et améliorer la précision du tracé.

Partie III : une justification du choix du modèle exponentiel obtenu dans le TP précédent, qui s'appuie sur des outils d'analyse, dérivée, primitive, intégration par parties, dans un contexte de probabilités et qui peut faire l'objet d'un TP.

B. Partie I

Il s'agit d'une activité qui peut servir de support au cours sur la fonction de répartition. En effet par le biais d'un exercice sur une variable aléatoire, il faut en fait dans un premier temps effectuer des calculs de probabilités d'évènements particuliers, puis la définition de la fonction de répartition étant donnée, réinvestir les calculs précédents avec cette fois-ci le vocabulaire particulier à ce nouvel objet.

Un joueur lance simultanément 3 *pièces de monnaie, parfaitement équilibrées. Il gagne* 60 F *s'il obtient* 3 *faces, gagne* 30 F *s'il obtient* 2 *faces, gagne* 10 F *s'il obtient* 1 *face, mais perd* 100 F *s'il n'obtient que des piles.*

X *est la variable aléatoire représentant le gain algébrique du joueur. Elle admet pour loi de probabilité :*

x_i	–100	10	30	60
p_i	1/8	3/8	3/8	1/8

Questions préliminaires :

∗ *Quelle est la probabilité de perdre au moins* 110 F *?*

P[X ≤ -110] = 0 puisque l'on ne peut pas perdre plus de 100 F, cet événement est impossible.

∗ *Quelle est la probabilité de perdre au moins* 10 F *?*

P[X ≤ -10] = P[X = -100] = 1/8.

∗ *Quelle est la probabilité de perdre de l'argent ?*

P[X ≤ 0] = P[X = -100] = 1/8.

∗ *Quelle est la probabilité que le gain soit inférieur ou égal à* 20 F *?*

C'est la probabilité de perdre 100 F ou de gagner 10 F :

P[X ≤ 20] = P[X = -100] + P[X = 10] = 1/2

∗ *Quelle est la probabilité que le gain soit inférieur ou égal à* 50 F *?*

C'est la probabilité de perdre 100 F ou de gagner 10 F ou de gagner 30 F :

P[X ≤ 50] =
P[X = -100] + P[X = 10] + P[X = 30] =
= 7/8

∗ *Quelle est la probabilité que le gain soit inférieur ou égal à* 1000 F *?*

C'est la probabilité de perdre 100 F ou de gagner 10 F, ou de gagner 30 F, ou de gagner 60 F. C'est la probabilité de l'événement certain :

P[X ≤ 1000] =
P[X = -100] + P[X = 10] + P[X = 30] + P[X = 60]
= 1

Définition :

On appelle fonction de répartition de X*, la fonction numérique* F *définie sur* **R** *par*

$$\forall\, x \in \mathbf{R},\ F(x) = P[\, X \leq x \,]$$

On peut remarquer que les questions préliminaires ne portaient en fait que sur des valeurs particulières de la fonction de répartition :

$P[X \leq -110] = F(-110) = 0$

$P[X \leq 0] = F(0) = p_1$

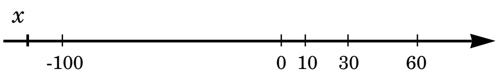
$P[X \leq 20] = F(20) = p_1 + p_2$

$P[X \leq 50] = F(50) = p_1 + p_2 + p_3$

$P[X \leq 1000] = F(100) = p_1 + p_2 + p_3 + p_4 = 1$

Détermination de la fonction de répartition :

Nous pouvons maintenant donner les valeurs de F(x) :



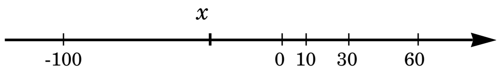
∗ *1er cas :* si x < -100

F(x) = P[X ≤ x] peut s'exprimer ainsi : *quelle est la probabilité que le gain soit inférieur ou égal à* x *francs où* x *est lui-même un nombre plus petit que* -100 *(perdre au moins* 200 *si* x = -200*)*.

On retrouve le cas particulier F(-110). L'événement [X ≤ x] est impossible puisque la perte ne peut dépasser 100 F.

Donc si $x \in]-\infty\,; -100[$, F(x) = 0 .

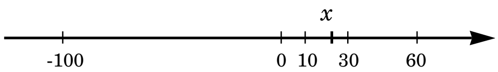
∗ *2ème cas :* si $-100 \leq x < 10$



F(x) = P[X ≤ x] *est la probabilité de gagner moins de* x *francs où* x *est un nombre compris entre* -100 *et* 10 *(perdre plus de* 90 F, 50 F, 1 F, *gagner moins de* 5 F *...).* Pour gagner moins de x francs, quand x est compris entre -100 et 10, on ne peut que perdre 100 F :

$F(x) = P[\, X = -100 \,] = p_1$, si $x \in [-100\,; 10[$.

∗ *3ème cas :* si $10 \leq x < 30$

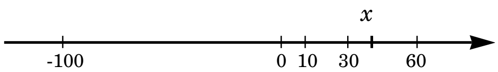


F(x) = P[X ≤ x] *est la probabilité de gagner moins de* x *francs où* x *est un nombre compris entre* 10 *et* 30. On retrouve le cas particulier F(20) :

$F(x) = P[\, X = -100 \,] + P[\, X = 10 \,] = p_1 + p_2$,

si $x \in [10\,; 30[$.

∗ *4ème cas :* si $30 \leq x < 60$

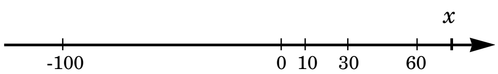


F(x) = P[X ≤ x], x *est un nombre compris entre* 30 *et* 60. On retrouve le cas particulier F(50) :

$F(x) = P[\, X = -100 \,] + P[\, X = 10 \,] + P[\, X = 30 \,]$
$= p_1 + p_2 + p_3$

si $x \in [30\,; 60[$.

∗ *5ème cas :* si $x \leq 60$

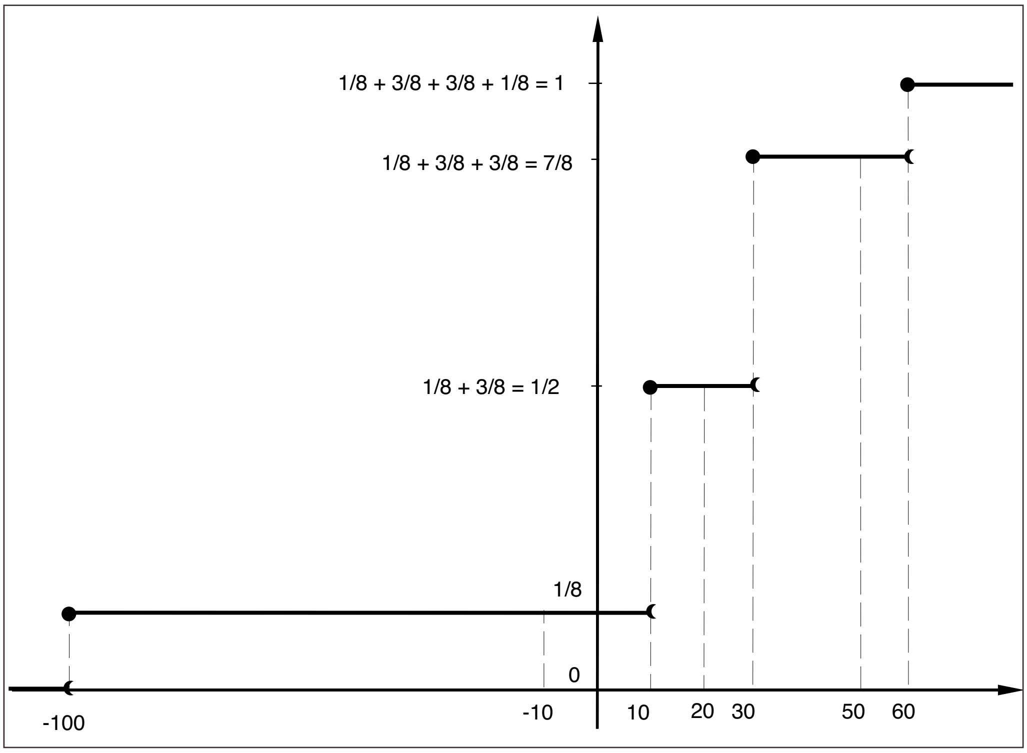


F(x) = P[X ≤ x], x *est un nombre plus grand que* 60. On retrouve le cas particulier F(1000). Dans ce cas, *tous les gains* sont envisageables :

$F(x) = P[\, X = -100 \,] + P[\, X = 10 \,] + P[\, X = 30 \,] + P[\, X = 60 \,] = p_1 + p_2 + p_3 + p_4 = 1$

si $x \in [30\,; +\infty[$.

Tracé de la fonction de répartition



Il ne reste plus qu'à tracer la courbe représentative de F.

Une fois le graphique tracé, il est intéressant de faire apparaître les différences avec les graphiques qu'ils ont tracés jusque-là : points de discontinuité, les sauts qui ne sont autres que les probabilités qu'ils ont calculées tout au début pour la loi de probabilité.

Nous sommes tout à fait conscients que cette présentation est souvent répétitive. Mais encore une fois, elle paraît importante pour une bonne compréhension de cette notion. Nous avons essayé de la présenter aux élèves de la façon la moins rébarbative possible, en utilisant un vocabulaire qu'ils maîtrisent bien.

C. Partie II : première proposition de TP

Ce TP est destiné à des élèves de Terminale, avec quelques variantes possibles suivant qu'il s'agit d'élèves de Terminale ES ou non. La fonction de répartition que nous allons rencontrer ici, contrairement à l'activité précédente, est continue.

Une entreprise veut étudier la fiabilité *d'un certain type de matériel. Pour cela elle prélève un échantillon de* n = 20 *éléments de sa production et contrôle toutes les* 500 *heures, le bon fonctionnement de ces éléments. On obtient ainsi le nombre d'éléments défaillants par intervalle de* 500 *heures.*

Tableau 1 :

Intervalle de temps en heures dans lequel intervient la première défaillance	Effectif
[0,500]	7
]500,1000]	4
]1000,1500]	3
]1500,2000]	2
]2000 ,2500]	2
]2500,3000]	1
]3000,4000]	1

Considérons tout d'abord le problème d'un point de vue statistique.

Le tableau précédent peut-être lu comme un tableau de données d'une série statistique discrète, définie de la manière suivante :

au bout de t_1 = 500 *heures, il y a* 7 *éléments défaillants,*

au bout de t_2 = 1000 *heures, il y a* 4 *éléments défaillants de plus,*

…

Avec les notations statistiques usuelles, nous obtenons le tableau suivant :

Tableau 2 :

temps t_i en heures	effectif n_i	fréquence f_i	fréquence cumulée
500	7	0.35	0.35
1000	4	0.20	0.55
1500	3	0.15	0.70
2000	2	0.10	0.80
2500	2	0.10	0.90
3000	1	0.05	0.95
4000	1	0.05	1

La fonction de répartition associée à cette série statistique est définie par :

$$\forall\, t \in \mathbf{R}\,,\ f(t) = \sum_{t_i \le t} mf_i\ .$$

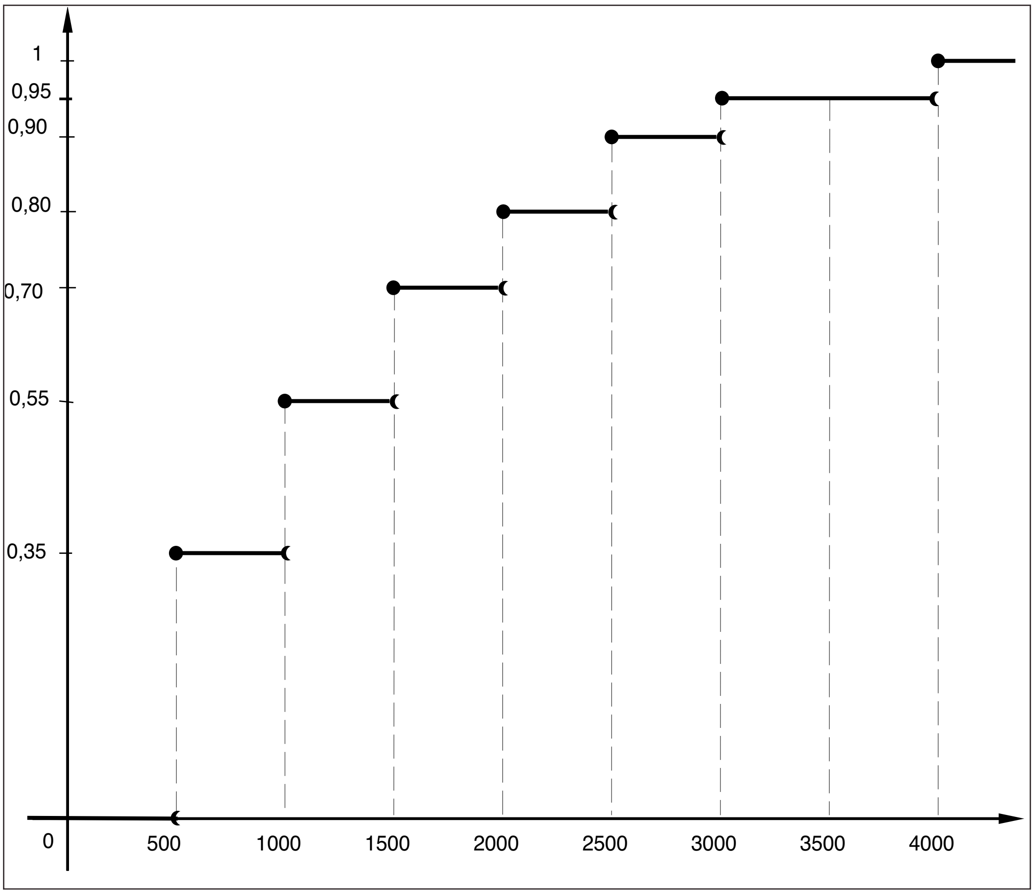
Elle admet la représentation graphique donnée sur la page suivante.

Cette représentation peut se lire ainsi : *avant* 500 *heures de fonctionnement, il n'y avait pas d'éléments défaillants ; entre* 500 *et* 1000 *heures, il y en a* 35 *% ; … ; après* 4000 *heures de fonctionnement, tous les éléments sont tombés en panne.*

L'analogie avec la fonction de répartition telle que nos élèves la connaissent saute ici aux yeux. Mais il faut reconnaître que dans cette situation, elle n'a pas beaucoup de sens. La première mesure étant effectuée au bout de 500 heures, si l'on se place du point de vue d'une série discrète, cela signifie qu'au bout de 500 heures exactement, 7 éléments tombent en panne en même temps.

Ceci paraît tout à fait invraisemblable. Il est donc préférable de supposer que les défaillances sont réparties uniformément dans chacun des intervalles.

Dans ce cas, le tableau 1 peut être lu cette



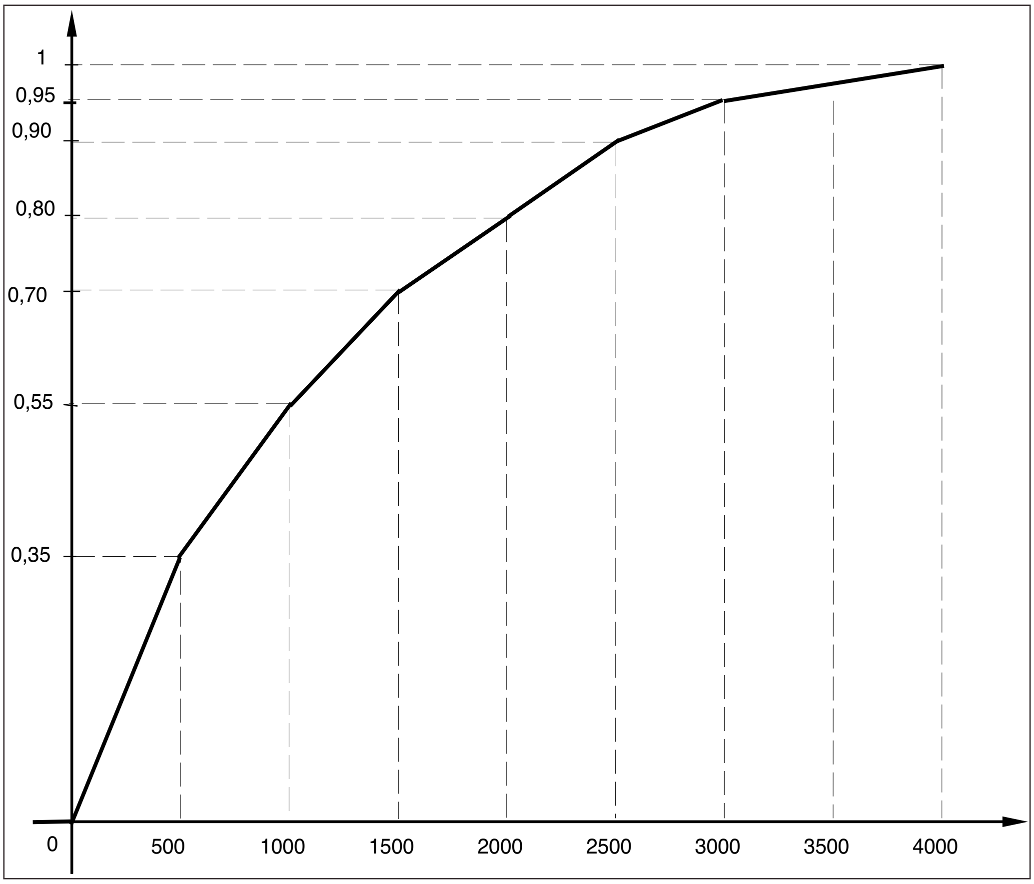
fois-ci comme les données d'une série statistique groupée. Nous pouvons alors tracer son polygone des fréquences cumulées, qui est aussi la représentation graphique de sa fonction de répartition. On obtient une fonction affine par morceaux. (cf. page suivante)

Nous sommes ainsi passés de la représentation graphique d'une fonction de répartition discrète à la représentation graphique d'une fonction de répartition continue.

Mathématisons le problème.

Désignons par T *la variable aléatoire qui à tout élément tiré au hasard dans une population constituée d'éléments du même type, associe son temps de bon fonctionnement, ou sa durée de vie avant une défaillance.*

$F(t)=P[\ T \leq t\ ]$ *est la probabilité qu'un dispositif choisi au hasard ait une défaillance avant l'instant* t (il n'est pas utile à ce niveau du TP de parler de fonction de répartition). Le tableau 1 ci-dessus se lit de la



manière suivante :

— au bout du temps t_1 = 500 heures, il y avait 7 éléments défaillants,

— au bout du temps t_2 = 1 000 heures, il y avait 7 + 4 = 11 éléments défaillants,

— au bout des 4 000 heures, les 20 éléments sont défaillants.

Comme il n'est pas possible de tester la population entière de ce type d'éléments, nous avons recours à un échantillon, choisi le moins mal possible, afin de tester la durée de vie de ces éléments. Et, bien que dans cet échantillon aucun élément n'ait une durée de vie supérieure à 4 000 heures, cela paraît tout à fait invraisemblable dans la population entière.

La méthode classique utilisée en statistique pour calculer une fréquence est celle que l'on a utilisée ci-dessus, c'est-à-dire $\frac{n_i}{n}$, qui est aussi utilisée dans les calculs de probabilité quand on parle de $\frac{\text{cas favorables}}{\text{cas possibles}}$.

Mais cette formule fait apparaître que

$F(4000) = \frac{20}{20} = 1.$

Il serait donc alors certain qu'au delà de 4000 heures, tous les éléments sont défaillants. Ce qui est en contradiction avec ce qui a été dit ci-dessus.

Il faut donc s'arranger pour avoir $F(4000) \neq 1$. Une méthode simple consiste à prendre :

$$F(t_i) = \frac{n_i}{n+1}.$$

Elle a l'avantage de ne pas trop modifier les résultats quand n = 20 (voir tableau 2 et l'encadré ci-dessous).

Avant de tracer la courbe représentative de F, il est indispensable de préciser le modèle avec les quatre points suivants :

— *On étudie* F *sur* [0 ; + ∞ [(on se limite à cet intervalle pour des raisons pratiques, si $x<0$, $F(x) = 0$). $F(0) = 0$, la courbe passe par l'origine

— F *est une fonction continue puisque la panne peut se produire à n'importe quel instant.*

— F *est une fonction croissante, la probabilité de tomber en panne croit avec le temps*

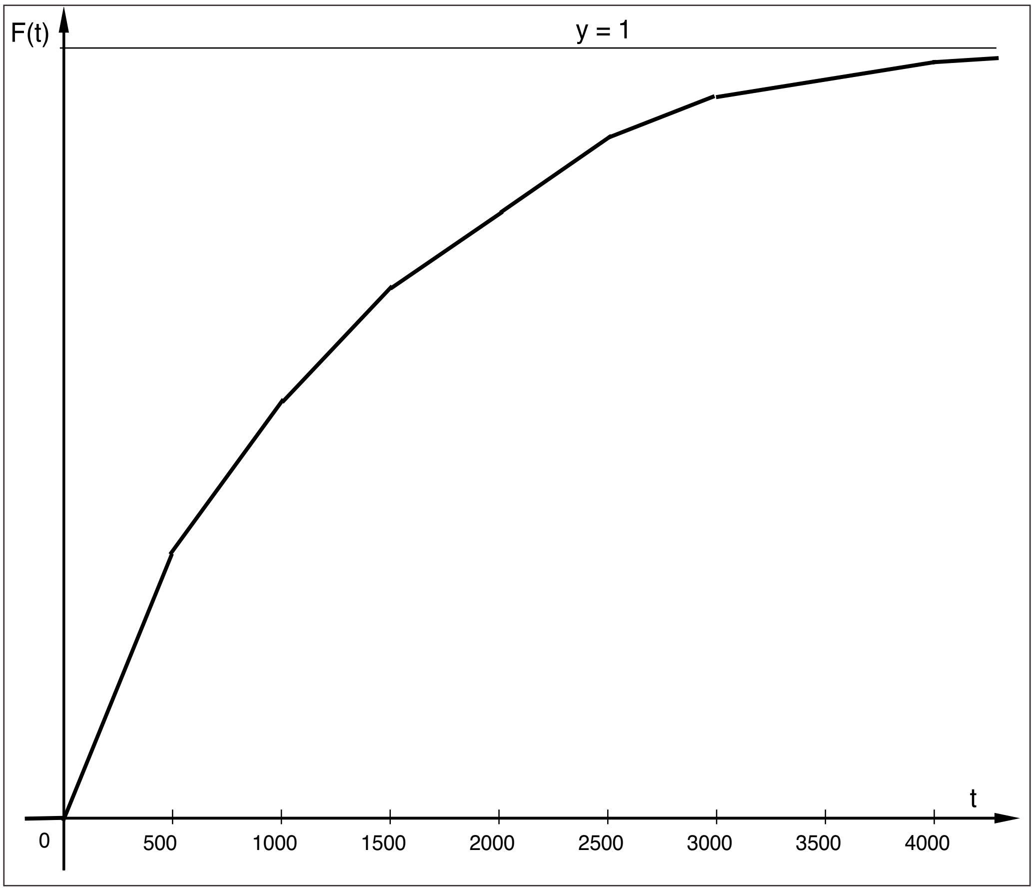
— $\lim_{t \to \infty} F(t) = 1$: *au bout d'un temps infini il est certain que tous les éléments seront défaillants, la droite d'équation* y = 1 *est asymptote à la courbe.*

A partir de ces renseignements, il peut être

Intervalle de temps	Nombre d'éléments défaillants	Instant t_i en heures	Nombre n_i d'éléments défaillants à l'instant t_i	Estimation de $F(t_i) = \frac{n_i}{n+1}$.
[0,500]	7	500	7	0.33
]500,1000]	4	1000	11	0.52
]1000,1500]	3	1500	14	0.67
]1500,2000]	2	2000	16	0.76
]2000 ,2500]	2	2500	18	0.86
]2500,3000]	1	3000	19	0.90
]3000,4000]	1	4000	20	0.95

Nous obtenons de cette manière une estimation de la fonction F, appelée fonction de défaillance, dont on connaît un tableau de valeurs :

t_i	0	500	1000	1500	2000	2500	3000	4000
$F(t_i)$	0	0.33	0.52	0.67	0.76	0.86	0.90	0.95



intéressant de laisser les élèves la tracer.

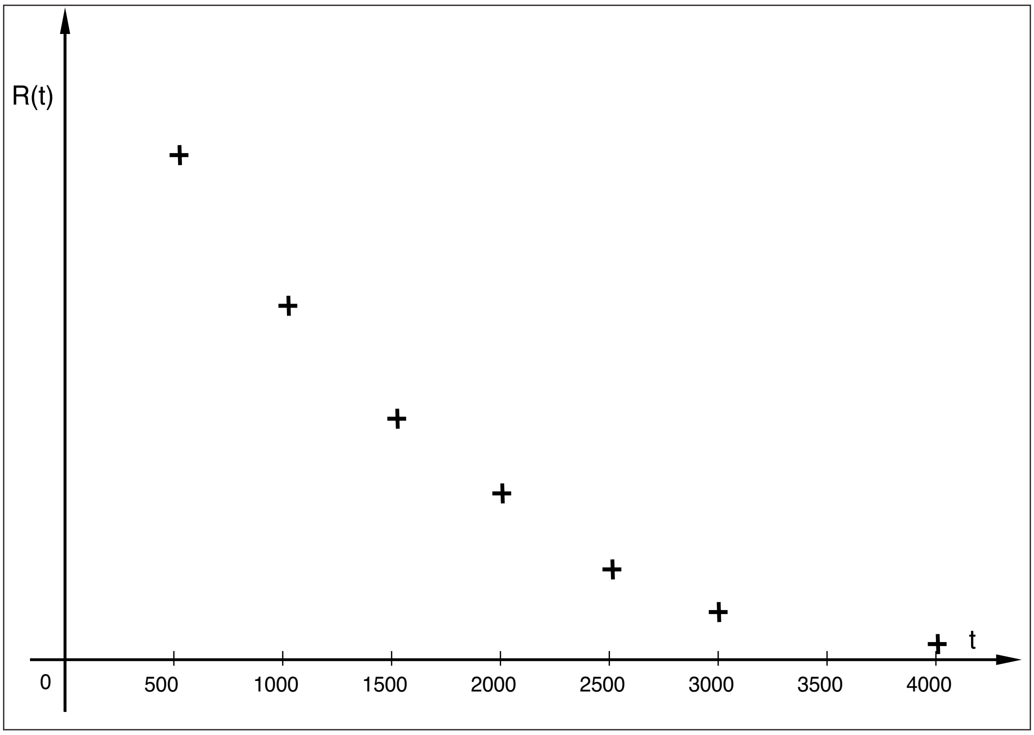
Afin de préciser la courbe représentative de F, nous allons voir qu'il est possible d'approfondir les constatations qui viennent d'être faites. *Posons :*

$R(t) = 1 - F(t) = 1 - P[\ T \leq t\ ] = P[\ T > t\ ]$.

R(t) *est la probabilité qu'un élément pris au hasard ait un temps de bon fonctionnement supérieur à un temps* t *donné.* R *est la fonction de fiabilité. A partir de l'estimation de* F *effectuée ci-dessus, nous obtenons l'estimation de* R *donnée ci-dessous.*

Nous allons désormais nous intéresser

t_i	500	1000	1500	2000	2500	3000	4000
$R(t_i)$	0.67	0.48	0.33	0.24	0.14	0.10	0.05



au nuage de points (t_i, R(t_i)).

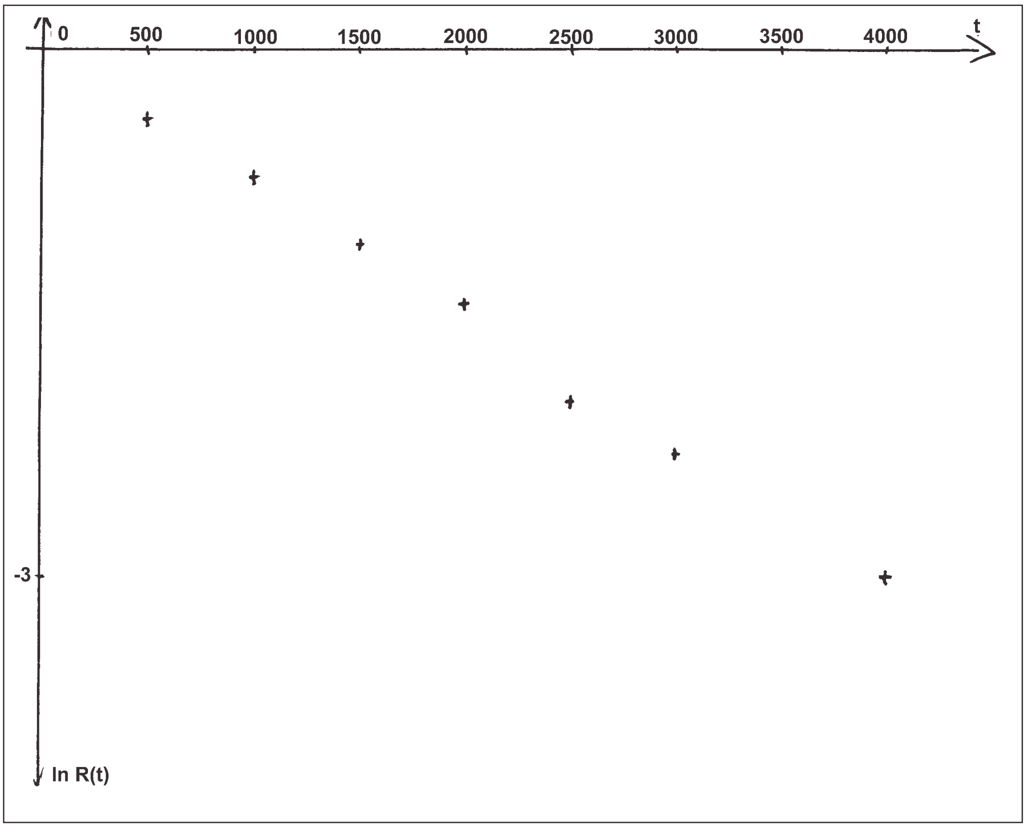
Ce nuage de points n'est manifestement pas «rectiligne». Etant donné la forme de ce nuage, regardons l'allure du nuage de point (t_i, ln R(t_i)). (Cf. la figure page suivante et le tableau ci-dessous).

La méthode utilisée peut différer suivant les classes dont nous avons la charge :
— pour les sections n'ayant pas d'outils statistiques, on trace le nuage de points (t_i, ln R(t_i)), on constate qu'il est «rectiligne» et à partir de deux points, on détermine une équation de droite approximant ce nuage.

— Pour la section ES : puisque le nuage de points (t_i, ln R(t_i)) est de forme rectiligne, c'est une occasion de réinvestir le travail effectué sur le papier semi-log et la méthode des moindres carrés.

L'équation de la droite ainsi trouvée est :

t_i	500	1000	1500	2000	2500	3000	4000
ln R(t_i)	-0.40	-0.73	-1.11	-1.43	-1.97	-2.30	-3



$$y = -7{,}6 \times 10^{-4} t + 0{,}0294$$

$$\text{soit } \ln R(t) = -7{,}6 \times 10^{-4} t + 0{,}0294$$

$$\text{donc } R(t) = e^{-7{,}6 \times 10^{-4} t + 0{,}0294}$$

$$\text{or } e^{0{,}0294} \approx 1{,}03$$

$$\text{donc } R(t) \approx e^{-0{,}00077t}$$

La courbe représentative de l'estimation de F est donc très proche de la courbe représentative de la fonction exponentielle :

$$t \mapsto 1 - e^{-0{,}00077t}$$

La première partie de ce TP nous avait permis, à partir d'un tableau de valeurs, d'effectuer un premier tracé de la courbe représentative de l'estimation de F. Comme il n'y avait que huit valeurs, le tracé était relativement imprécis. La fonction exponentielle ci-dessus (qui remplit les quatre conditions évoquées précédemment) nous permet de tracer une nouvelle courbe représentative de l'estimation de F.

Nous avons donc maintenant à notre disposition quatre représentations graphiques de l'estimation de F : deux obtenues à partir de modèles statistiques, deux obtenues à partir d'un modèle probabiliste. Il peut être intéressant alors de faire comparer les deux courbes représentatives aux élèves, en leur précisant bien que ce sont ces courbes qui sont utilisées pour obtenir des valeurs de F(t) et de R(t), afin d'estimer la fiabilité des composants.

C'est à ce moment qu'il nous paraît judicieux de faire remarquer que F est une fonction de répartition, telle qu'ils l'ont énoncée dans le cours de probabilités, puisque F a été définie par $F(t) = P[\,T \leq t\,]$. Et qu'il existe des fonctions de répartition qui ne sont pas des fonctions en escalier.

D. Partie III : une justification mathématique

Dans la partie II, nous avons amené les élèves à modéliser le phénomène étudié par le modèle exponentiel. Essayons de justifier le choix de ce modèle.

Le problème que l'on se pose ici est un peu différent :

Quelle est la probabilité qu'un élément tombe en panne entre les instants t *et* t + h*, sachant qu'il a fonctionné jusque-là.*

Il s'agit d'une probabilité conditionnelle :

$$P\left({}^{[\,t < T \leq t+h\,]}/_{[\,T > t\,]} \right) =$$

$$= \frac{P(\,[\,t < T \leq t+h\,] \cap [\,T > t\,]\,)}{P(\,[\,T > t\,])}$$

or $[\,t < T \leq t+h\,] \cap [\,T > t\,] = [\,t < T \leq t+h\,]$, ceci revient donc à calculer :

$$\frac{P(\,[\,t < T \leq t+h\,]\,)}{P(\,[\,T > t\,])}$$

mais $P(\,[\,t < T \leq t+h\,]\,) = F(\,t+h\,) - F(\,t\,)$ et $P(\,[\,T > t\,]\,) = R(\,t\,)$. La probabilité cherchée s'écrit finalement :

$$\frac{F(\,t+h\,) - F(\,t\,)}{R(\,t\,)}$$

Intéressons nous au nombre :

$$\lim_{n \to \infty} \frac{F(\,t+h\,) - F(\,t\,)}{h.R(\,t\,)}$$

appelé *taux d'avarie instantané* à l'instant t et noté $\lambda(t)$.
Nous remarquons que :

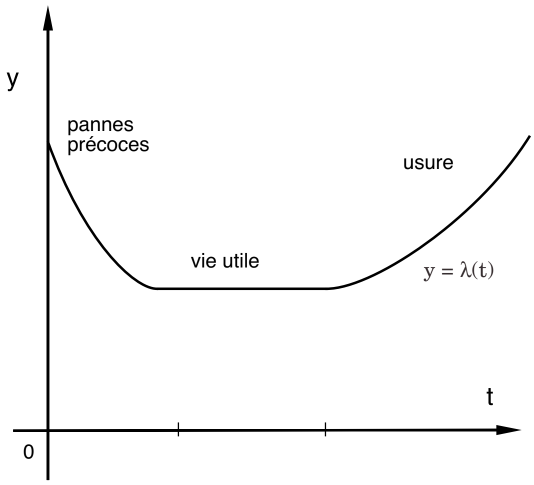
$$\lim_{n \to \infty} \frac{F(\,t+h\,) - F(\,t\,)}{h} = F'(\,t\,),$$

$\lambda(t)$ peut donc s'écrire :

$$\lambda(t) = \frac{F'(\,t\,)}{R(\,t\,)} = \frac{F'(\,t\,)}{1 - F(\,t\,)}.$$

Expérimentalement, on constate que pour la plupart des matériels, la courbe représentative de la fonction $t \longmapsto \lambda(t)$ a la forme d'une courbe dite courbe en baignoire.

La période dite de «vie utile» correspond



à $\lambda(t)$ à peu près constant.

Reprenons l'expérience qui nous a servi de fil conducteur tout au long du TP précédent.

Nous allons nous intéresser au rapport :

$$\frac{F(t+h)-F(t)}{h.(1-F(t))},$$

pour les valeurs de F(t) que nous connaissons ; h vaudra 500 puisque les mesures sont effectuées toutes les 500 heures :

$$\frac{F(500)-F(0)}{500(1-F(0))} \approx 6{,}6 \times 10^{-4}$$

$$\frac{F(1000)-F(500)}{500(1-F(500))} \approx 5{,}7 \times 10^{-4}$$

$$\frac{F(1500)-F(1000)}{500(1-F(1000))} \approx 6{,}3 \times 10^{-4}$$

$$\frac{F(2000)-F(1500)}{500(1-F(1500))} \approx 5{,}5 \times 10^{-4}$$

$$\frac{F(2500)-F(2000)}{500(1-F(2000))} \approx 8{,}3 \times 10^{-4}$$

$$\frac{F(3000)-F(2500)}{500(1-F(2500))} \approx 5{,}7 \times 10^{-4}$$

$$\frac{F(4000)-F(3000)}{1000(1-F(3000))} \approx 5 \times 10^{-4},$$

[ici *h* vaut 1000 puisque la dernière mesure a été effectuée au bout de 1000 heures].

Nous pouvons remarquer que les valeurs trouvées sont très proches les unes des autres. Ceci laisse supposer que le taux d'avarie instantané est à peu près constant. Nous pouvons donc supposer que les éléments testés étaient dans leur période de vie utile. Nous sommes en présence de ce que l'on appelle le modèle exponentiel. Pour poursuivre les calculs, nous prendrons comme valeur de λ la moyenne des valeurs précédemment trouvées, c'est-à-dire $6{,}2 \times 10^{-4}$. Dans le TP précédent, nous avions trouvé en approximant le nuage de points par la méthode des moindres carrés $7{,}6 \times 10^{-4}$ comme valeur de λ. L'écart entre ces deux nombres s'explique par la qualité des approximations utilisées. La méthode des moindres carrés est ici *a priori* beaucoup plus fiable que les calculs « à la main » des valeurs de $\lambda(t)$.

Pour que ces dernières soient plus proches de la réalité, il aurait fallu disposer de mesures plus rapprochées, c'est-à-dire de valeurs de λ plus petites puisque λ est censé tendre vers 0, ce qui est loin d'être le cas lorsque t = 500. La partie II et la partie III sont indépendantes, c'est pourquoi nous n'avons pas repris, telle qu'elle, l'approximation de λ qui avait été trouvée.

Comme $\lambda(t) = \dfrac{F'(t)}{1-F(t)}$; $0{,}00062 = \dfrac{F'(t)}{1-F(t)}$

Par intégration, nous obtenons :

$\ln(1 - F(t)) = -0{,}00062\, t + k$; $(0 \leq F(t) \leq 1)$

et comme F(0) = 0 ; on a donc k = 0 et

$1 - F(t) = e^{-0{,}00062t}$; $F(t) = 1 - e^{-0{,}00062t}$.

Dans la 2ème partie, nous avions trouvé comme approximation de la fonction F, par la méthode des moindres carrés, la fonction exponentielle : $t \longmapsto 1 - e^{-0{,}00062t}$, qui est relativement proche de la fonction trouvée ci-dessus.

Afin de réinvestir les outils d'analyse dans un contexte différent, nous pouvons poursuivre nos investigations. Il est intéressant de calculer le temps moyen de bon fonctionnement noté MTBF de part l'appellation anglo-saxonne (*mean time between failures*), qui est en fait :

$$\text{MTBF} = \text{E(T)} = \int_0^{+\infty} \text{t.F(t) dt} \,,$$

espérance de la variable aléatoire T.

$$\text{Posons : } \text{I}(\alpha) = \int_0^{\alpha} \text{t.F'(t) dt} \,,$$

$$\text{il vient : } \text{I}(\alpha) = \int_0^{\alpha} 0{,}00062.\text{t.e}^{-0{,}00062\text{t}} \,\text{dt} \,.$$

A l'aide d'une intégration par parties, nous obtenons :

$$\text{I}(\alpha) = -\,\alpha.\text{e}^{-0{,}00062\alpha} - \frac{\text{e}^{-0{,}00062\alpha}}{0{,}00062} + \frac{1}{0{,}00062}$$

$$\lim_{\alpha \to \infty} \text{I}(\alpha) = \frac{1}{0{,}00062} \approx 1612 \,.$$

Donc MTBF = 1612 h. On peut estimer que le temps moyen de bon fonctionnement d'un tel élément est de 1612 heures.

De manière générale, lorsque nous sommes en présence d'un modèle exponentiel, c'est-à-dire que $\text{R(t)} = \text{e}^{-\lambda\text{t}}$ ou $\text{F(t)} = 1 - \text{e}^{-\lambda\text{t}}$ (ici $\lambda = 0{,}00062$), le MTBF a pour valeur $1 / \lambda$. Nous allons utiliser cette propriété pour retrouver le MTBF graphiquement.

Si $\text{t} = 1/\lambda$, $\text{R(t)} = \text{e}^{-\lambda \times (1/\lambda)} = \text{e}^{-1} \approx 0{,}368$. Donc

pour $\lambda = \dfrac{1}{0{,}00062}$, $\text{R(t)} \approx 0{,}368$.

L'abscisse du point de la droite (qui nous a servi d'approximation du nuage de points) *d'ordonnée* ln 0,368 *nous donne le temps moyen de bon fonctionnement.*

Il nous a semblé intéressant, dans un contexte de probabilités, où les outils mathématiques utilisés sont relativement restreints (en ce qui concerne les classes de Terminales), de voir fonctionner les notions les plus usuelles d'analyse : limite, dérivée, fonction exponentielle, intégration ... Et nous pensons que cela peut faire l'objet d'un TP qui soit un prolongement du précédent puisque l'on y retrouve la fonction de répartition F. Mais ces deux TP seraient indépendants. Alors que le premier TP utilisait très peu de connaissances mathématiques et portait davantage sur les courbes représentatives, ce deuxième TP plus technique ferait appel à de nombreux outils d'analyse. Il pourrait donc plutôt être proposé à des élèves de Terminale S.

E. Conclusion

Dans un premier temps cet article a tenté de donner du sens à une notion difficile à présenter en classe de Terminale : la fonction de répartition. Difficile pour deux raisons : la rencontre des fonctions en escalier, donc de fonctions non continues d'une part, et le côté abstrait de cette notion d'autre part.

Dans un deuxième temps, il s'appuie sur cet outil pour amener les élèves à un des actes les plus difficiles : la modélisation. C'est pourquoi il nous a paru important de l'illustrer par un phénomène concret. L'objectif est en même temps de préparer les élèves à une utilisation différente des probabilités dans les classes post-bac.

Le troisième temps est bien entendu la justification mathématique du modèle.

Les professeurs de mathématiques amenés à présenter la fiabilité partageront sûrement notre avis, à savoir que c'est un des chapitres les plus motivants à enseigner. Les autres verront dans la démarche proposée la mise en œuvre d'une logique scientifique : expérimentation, modélisation, justification.

Le choix de la fiabilité apporte un intérêt supplémentaire quand on sait l'importance de cette notion dans notre société. Ceux qui ont assisté à la conférence de Marc J. Pélegrin, lors des journées nationales de l'APMEP à Albi en 1996, se souviendront sûrement de ce passage. Il existe trois niveaux de panne : non critique, critique et cruciale. La panne critique correspond à une probabilité d'apparition de 10^{-7} par heure de vol. Le danger est potentiel mais contrôlable à court terme par l'équipage, il faut atterrir sur le terrain le plus proche. Nous ne parlerons pas ici de la panne cruciale. Marc J. Pélegrin soulignait qu'il est difficile de justifier une fiabilité meilleure que 10^{-3} / h par une expérimentation. Pour démontrer expérimentalement que la probabilité d'un accident catastrophique d'un système est inférieure à 10^{-7} (avec une confiance de 0,9), il faut effectuer 2625 ans d'essais sans rencontrer de pannes !

Ce travail sera nettement mis en valeur, si nous pouvons bénéficier de vos suggestions et comptes-rendus d'expériences concernant les activités présentées dans cet article.

e-mail : ppombour@ac-toulouse.fr

ELEMENTS DE BIBLIOGRAPHIE

SAINT-PIERRE G. et VERLANT B. : *Statistique et probabilités*, Foucher, 1993.

FAURE P., ASTIER J.-D., BOUCHON B. : *Mathématiques BTS industriel*, Nathan, 1992.

LEBOEUF C., ROQUE J.-L., GUEGAND J. : *Cours de probabilités et de statistiques* - deuxième édition, Ellipse, 1987.

BIGOT B., CHAPUT B., DANIEL J.-C., DUPERRET J.-C. : *Probabilités et statistiques, statistique inférentielle (BTS)*, Université de REIMS et Irem de REIMS, 1996.

PELEGRIN M. : «Les exigences de la sécurité aéronautique», *APMEP bulletin n°410,* 1996.