

Mth 2302 D

TP 5

Fiabilité & Régresser

Marc Bourdeau

ACCÉLÉRER LE VIEILLISSEMENT POUR ESTIMER UNE FIABILITÉ

*What is it like growing old Mr Fields ?
— It's better than the alternative.*

W.C. Fields [1880–1946],
comédien, et « *ad libber* » étatsunien.

Déterminer la fiabilité d'un produit, soit en cours de mise au point, soit après modification, est une étape essentielle avant la mise en marché de la vaste majorité des produits technologiques. Ce devoir, construit essentiellement sur un exemple, illustre la démarche type qu'on trouve en étude de la fiabilité pour vérifier la conformité à des spécifications du cahier des charges.

Bien des raffinements ont été développés ces dernières années, notamment dans l'estimation des paramètres de la loi de la durée de vie, mais la démarche exposée ici est encore essentiellement celle utilisée.

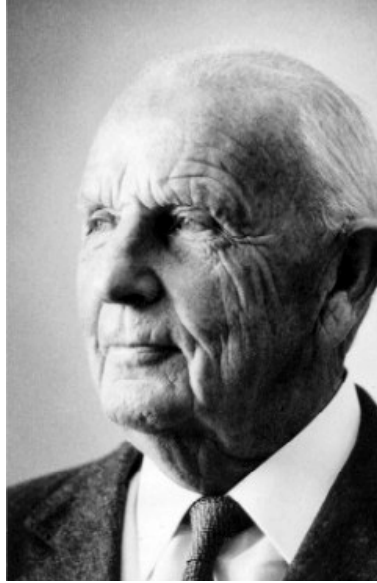


FIG. 1 – Portrait de [Ernst Hjalmar Wallodi Weibull](#) [1887–1979] par le photographe Sam C. Saunders. On place ce TP sous la figure tutélaire de ce brillant ingénieur.

La fiabilité

La fiabilité d'un « outil » au temps t , $R(t)$, est définie comme sa probabilité de fonctionnement au temps t . On se donne implicitement la variable aléatoire T , défini pour les valeurs non négatives de la droite réelle. T est « le temps à la première panne » de l'outil, on dit aussi sa loi de la durée de vie. Ainsi $R(t) \equiv P[T \geq t] = 1 - F_T(t)$, où $F_T(t)$ est la cumulative¹ de la loi T . La moyenne de cette VA est dite la *moyenne du temps de bon fonctionnement*, la MTBF. Plusieurs modèles sont utilisés pour T , notamment la loi exponentielle, la loi log-gaussienne, mais il est devenu presque universel d'utiliser la loi de Weibull².

¹ F_T étant par convention universelle la notation pour la cumulative d'une loi de probabilité ou variable aléatoire T , on ne peut l'utiliser pour la fiabilité. On utilise alors $R_T(t)$, ou simplement $R(t)$, de sa notation anglaise pour *reliability*. Dans d'autres contextes, on utilise S_T , pour la probabilité de *survie* au temps t .

²Wallodi Weibull, 1887-1979, ingénieur suédois bien connu pour ses travaux en fatigue

Pour un grand nombre d'outils, le fabricant doit garantir une certaine fiabilité à un temps prescrit. Par exemple, on doit garantir une MTBF de tant d'années. Bien évidemment, cela se traduit aussi en termes de probabilité du survie à un temps donné. Par exemple, on doit garantir que 95% des outils seront encore en fonctionnement à un temps prescrit, *i.e.* que seulement 5% des pannes se produiront avant ce temps. La plupart du temps, il est impossible pour les fabricants de tester ces garanties tant les MTBF sont grandes. Un exemple fera voir pourquoi.

1. On veut qu'une mission spatiale devant durer 10 jours ait seulement une chance sur 100 d'avoir une défaillance. On doit donc avoir une fiabilité de 99% à 10 jours.
 - (a) Supposons que le temps de bon fonctionnement suive une loi exponentielle. Trouver la MTBF minimale pour cet « outil ». *Indice* : Calculer la fiabilité $R_T(t) = P[T \geq t]$, et raisonner à partir de celle-ci. On comptera en heures.
 - (b) Supposons maintenant, pour simplifier, que l'engin spatial est composé de $N = 1\,000$ composants en série (ce nombre est de loin inférieur à la réalité). Supposons aussi que l'hypothèse d'indépendance des lois de durées de vie de chacun est avérée, et que celles-ci sont des lois exponentielles identiques.
 Quelle est la MTBF nécessaire pour chacun d'entre eux pour que l'outil dans son ensemble ait la fiabilité désirée ? On raisonnera en général pour un N quelconque. Puis on posera $N = 1\,000$. On exprimera la MTBF en heures.

Dans ces circonstances, on peut penser, pour accéder aux paramètres de la fiabilité, tester les outils en grand nombre, mais, encore là, ces nombres sont impraticables, ne fût-ce qu'à cause des coûts prohibitifs. On a imaginé cependant accélérer le vieillissement des outils en augmentant les contraintes du fonctionnement normal. Le vieillissement des composants et outils est le plus souvent dû à des propriétés physico-chimiques qui se détériorent avec le temps, c'est la *fatigue* des matériaux. Si on réussit à accélérer le vieillissement par augmentation des contraintes, sans pourtant changer les modes de défaillances, et qu'on peut extrapoler les propriétés du vieillissement aux

des matériaux. C'est dans son article : *A statistical distribution of wide applicability*, Jour. of Applied Mech. 18, 1951, 293-297, qu'il a présenté cette loi qu'on applique aujourd'hui partout en étude de la fiabilité. On pourra avec profit consulter le manuel sur cette loi ; une page sous le site suivant pour des éléments de biographie : <http://www.barringer1.com> ; enfin les sites <http://www.barringer1.com> ou <http://www.weibull.com> pour des détails techniques sur la fiabilité.

contraintes normales d'usage, on pourra chiffrer les paramètres, et assurer des garanties.

On comprend aisément que les propriétés physico-chimiques des matériaux sont altérées par la température, c'est en fait d'ailleurs la principale cause de la fatigue de beaucoup de matériaux. Dans les composants électroniques, on utilise beaucoup les voltages et les courants pour augmenter les températures de fonctionnement au delà de celles en utilisation normale, et on peut extrapoler les conclusions obtenues sur la fiabilité à l'aide la loi d'Arrhenius qui relie la vitesse d'une réaction chimique à sa température³.

Dans ce qui suit, on présente sur un exemple la démarche statistique des essais accélérés de fiabilité.



La démarche

On veut estimer les paramètres de la loi du bon fonctionnement d'un appareil, sa durée de vie utile. Soit T , cette variable aléatoire, on dit aussi la le temps à la (première) panne. Une fois cela connu, on peut en tirer les valeurs des quantiles T_α . Par exemple : quelle est le temps où en moyenne 95% des appareils sont encore en fonctionnement, soit $t_{0,05}$.

Il s'agit essentiellement d'accélérer le processus de vieillissement des appareils, en utilisant une variable contrôlée. Ici ce sera la température de fonctionnement⁴.

1. Un certain nombre d'appareils sont mis en marche, avec des plus hautes températures de fonctionnement que le fonctionnement habituel; ils vieillissent de façon accélérée. On utilise ici 3 températures différentes. Sont mis sur banc d'essai un nombre déterminé d'appareils à chacune des trois températures.

³On trouve en français un excellent petit livre d'introduction à la fiabilité : Pierre Chapouille, *La fiabilité*, Presses universitaires de France, coll. Que-sais-je ?, N° 1480, Paris, 1972.

⁴Tout un chacun a pu constater, par exemple, que l'exposition au soleil fait vieillir prématurément les peaux humaines. Dans ce cas, c'est le temps d'exposition au soleil qui est la variable accélérante.

2. Pour chacune des températures de fonctionnement, on obtient les paramètres d'un modèle pour les temps de bon fonctionnement, en l'occurrence des lois de Weibull.

Ces paramètres permettent d'obtenir des estimations des quantiles $T = t_\alpha$ pour le loi du bon fonctionnement à ces températures.

3. Pour passer aux paramètres des lois de bon fonctionnement aux températures habituelles de fonctionnement, on utilise la loi d'Arrhenius qui permet d'extrapoler des températures où le vieillissement est accéléré aux températures d'usage.

Nous allons modéliser la fiabilité par des lois de Weibull; d'autres lois sont aussi utilisées, mais la Weibull est la plus fréquente. Nous procéderons d'abord à une linéarisation de celle-ci pour en trouver aux moindres carrés les paramètres. Ceux-ci permettent d'en estimer les quantiles, i.e. les probabilités de survie exprimée par les quantiles : ce sont ces quantités qui sont d'intérêt.

Il peut être utile pour bien comprendre l'ajustement aux moindres carrés de faire quelques expériences des propriétés des modèles linéaires à l'aide des animations suivantes (cliquer sur les icones marginaux).



Accélérer le vieillissement

La compagnie N*** veut s'assurer que ses condensateurs électrolytiques fonctionnent au moins 10 ans dans les conditions normales d'utilisation, soit la durée de la garantie. Pour être précis, on veut garantir que le temps T de bon fonctionnement des condensateurs soit tel que $T_{0,05} \geq 10$ ans. On suppose que ces condensateurs sont utilisés de façon continue. Leur durée de vie est très sensible à la chaleur du fonctionnement qui dépend de la présence d'un courant dit de *ronflement*. Celui-ci est normalement de quelques dizaines de milliampères, auquel cas on trouve une température de fonctionnement moyenne de 50°C. Le courant de ronflement est facilement contrôlable, et c'est celui-ci qu'on utilise pour déterminer les valeurs de la variable accélérante, soit la température de fonctionnement.

On admet que le temps T de bon fonctionnement suit une loi de Weibull. La cumulative de cette loi est $F(t) = 1 - e^{-(\frac{t}{\theta})^\beta}$. β est dit le paramètre de *forme*, et θ est dit le paramètre d'*échelle* de la loi de Weibull⁵.

Aux fins de l'observation et de la détermination des paramètres par la méthode des moindres carrés, on linéarise cette cumulative et on obtient :

$$\ln\left[\ln\left(\frac{1}{1-F(t)}\right)\right] = \beta \ln(t) - \beta \ln(\theta). \quad (1)$$

1. Montrez l'équation (1).

Supposons qu'on ait m condensateurs en essai à une température donnée, et qu'on ait observé n « pannes » ou mauvais fonctionnements. Posons $t_{(1)} \dots, t_{(n)}$, les temps observés ordonnés des n pannes. Ces observations sont des approximations des quantiles $p_i = \frac{i}{m+1}, i = 1, \dots, n$ de $T : F(t_{(i)}) \approx p_i$. Ainsi donc, en vertu de l'équation (1), les points :

$$\left(\ln(t_{(i)}), \ln\left[\ln\left(\frac{1}{1-p_i}\right)\right] \right), \quad i = 1, \dots, n,$$

sont situés approximativement sur une droite $b_0 + b_1 \ln(t)$, dont les paramètres b_0 et b_1 servent à estimer les paramètres β et θ .

2. Trouver les expressions mathématiques pour β et θ en fonction des paramètres b_0 et b_1 de la régression.
3. Utiliser l'expression de la cumulative de la loi de Weibull pour calculer les quantiles t_p , *i.e.* les temps tels que $F(t_p) = p$ en fonction de β et θ , pour p dans l'intervalle $(0; 1)$.

Il faut accélérer le processus de vieillissement car il est impossible de tester les unités pendant de nombreuses décennies avant de les mettre en production, temps qu'il faudrait pour avoir une idée en temps réel de leur durabilité! On dispose en fait d'une année et de 300 unités pour tester les paramètres de durabilité. Après quoi, on avisera des améliorations à mettre au point avant de lancer la production à plus grande échelle.

On effectuera des essais à des températures *judicieusement* choisies et supérieures à celle du fonctionnement normal. Il convient en effet de choisir des températures qui ne seront pas stressantes au point de changer les *modes de défaillance*. Ainsi à de trop hautes températures, les matériaux pourraient fondre, et de telles défaillances ne se produiraient jamais en usage normal.

⁵On remarque que pour $\beta = 1$, on a la loi exponentielle.

L'allocation des 300 unités aux diverses températures choisies pour les essais se font selon des principes statistiques qu'il serait trop long d'exposer ici. Intuitivement, on conçoit qu'à des températures plus près des conditions normales, on testera plus d'unités qu'à de plus hautes températures, en partie parce que les défaillances sont plus rares⁶.

Plus loin (section), on expliquera comment il est possible d'extrapoler à des températures plus basses les résultats obtenus à de plus hautes températures, pour déterminer la probabilité de « survie » à 10 ans d'un certain pourcentage (ici 95%) des unités en usage : $P[T_{0,05} \geq 10 \text{ ans}]$.

Dans un premier temps donc, on met sur banc d'essais les 300 unités dûment vérifiés de la façon suivante : 171 unités sont testées à 78°C, 87 à 98°, et 42 à 120°. Les courants de ronflement permettent le contrôle exact des températures des condensateurs. On vérifie quotidiennement les fonctionnements des unités. Au bout d'une année, l'expérience est arrêtée.

On trouvera les données en cliquant sur l'icone marginal. Les trois variables de température sont notées $TEMP_i$; les trois variables des temps pour les unités qui ont failli : T_i , i étant 1, 2 ou 3, selon la température. Les temps T_i sont en jours et rapportent en fait les jours où on a constaté le mauvais fonctionnement des condensateurs.



Data

4. Pourquoi devrait-on assigner les unités expérimentales à chaque température de façon aléatoire? On appelle *randomisation* cette procédure, elle est fortement suggérée pour toute expérimentation statistique.
5. À l'observation des données, vous semble-t-il que les régressions linéaires sont appropriées dans chaque cas de température?
6. Après avoir tamisé les données pour ne garder que les données qui ne sont pas suspectes, obtenir les résultats des régressions (1) pour chaque température⁷.
7. Pour trouver un intervalle de confiance simultané de niveau $(1-\alpha)$ pour deux paramètres θ_i , $i = 1, 2$ à partir de deux estimateurs indépendants $\hat{\theta}_i$, $i = 1, 2$ des deux paramètres, il suffit de déterminer un intervalle de confiance de niveau $(1-\alpha)^{\frac{1}{2}}$ pour chacun d'entre eux. Donner et

⁶On trouvera des éléments de ces techniques d'allocation à la référence suivante : William Q. Meeker & Gerald J. Hahn, *How to plan an accelerated life test — Some practical guidelines*, American Society for Quality Press, Milwaukee WI, 1985.

⁷Attention : ce jeu de donnée tamisé sera celui en usage pour la suite du devoir. Et le tamisage, si on supprime des données, forcera un changement dans les quantiles. On ne se donnera la peine que de supprimer les données les plus aberrantes, les autres incongruités ne seront pas traitées : il faut aller de l'avant...

justifier le résultat analogue pour un intervalle de confiance simultané de k paramètres.

8. Calculer un intervalle de confiance simultané⁸ de niveau 95% pour les paramètres b_1 et b_0 de la régression à 98°C.
9. Tirer du calcul précédent une fourchette à 95% pour la valeur de θ à 98°C.

Note : on devra raisonner assez finement ici, car l'estimateur pour θ est une fonction des deux estimateurs \hat{b}_0, \hat{b}_1 , d'où la nécessité d'estimateurs simultanés pour ces derniers. De plus, les considérations mathématiques impliquées feront voir que le calcul numérique est très sensible à la précision des calculs.



Estimations aux conditions d'usage

La loi d'Arrhenius (Svante Arrhenius⁹, physicien suédois 1859-1927, prix Nobel de chimie 1903), permet d'extrapoler les résultats obtenus à de plus hautes températures à la température de fonctionnement de 50°C. Cette loi relie en effet la température où s'effectue une réaction à sa vitesse. Soit T le temps de bon fonctionnement (de survie) des unités $t_{0,05}$, son 5^e centile à une température donnée (les quantiles sont trouvés expérimentalement à l'aide de la formule explicitée plus haut). Alors le modèle suivant de régression simple s'applique à l'extrapolation pour d'autres températures :

$$y = \beta_0 + \beta_1 x + \varepsilon$$

avec $y \equiv \ln(t_{0,05})$ à une température donnée et $x \equiv 1/(\text{°C} + 273)$, soit l'inverse de la température en degrés Kelvin.

⁸On sait que l'hypothèse d'indépendance des deux paramètres b_0, b_1 n'est pas vérifiée pour une régression, mais, pour les intervalles de confiance simultanés, on peut montrer que cela ne change pas grand chose en pratique pourvu que $1 - \alpha$ soit assez grand.

⁹Un autre physicien et astronome suédois est bien connu, il s'agit de Anders Celsius, 1701-1744, créateur de l'échelle éponyme des températures, qui joue aussi un rôle ici... Décidément ce devoir fut rédigé sous des auspices suédois.

On dispose donc ici par calcul, grâce aux estimations de la section précédente et à la formule obtenue au numéro 3, d'une estimation de y à chacune des trois températures de l'expérimentation.

10. En utilisant les paramètres donnés par la régression construite sur le modèle d'Arrhenius, donner un intervalle de confiance à 95% pour une *nouvelle* observation de la valeur de $\ln(t_{0,05})$ à 98°C. En déduire un intervalle de confiance de niveau 95% pour $t_{0,05}$ à cette température. Commenter.

Note : vous commenterez ici aussi sur la sensibilité et la précision exigée des calculs.

11. Estimer la valeur de $\ln(t_{0,05})$ à 50°. Déduire de cette valeur, ainsi que des propriétés statistiques des *prédictions* issues des régressions, la proportion estimée des condensateurs des unités produites actuellement telle que $t_{0,05}$ à 50° dépasse dix ans. Vous terminerez en commentant sur la valeur de la garantie.

Note : Pour cela, on aura besoin de calculer l'écart type estimé de $\ln(t_{0,05})$ à 50°. Celui-ci ne se calcule pas aisément par sa formule explicite, mais s'obtient aisément en remarquant que l'étendue de l'intervalle de prédiction de niveau $1 - \alpha$, calculé automatiquement par Statistica lorsqu'on cherche une prédiction, vaut $2T_{\nu, 1-\frac{\alpha}{2}} \hat{\sigma}_p$ où $\hat{\sigma}_p$ est justement l'écart type estimé d'une prédiction.



Validation des résultats

En réalité, comme les examens des unités sur les bancs d'essai se font tous les jours, les pannes constatées un jour donné ont pu se produire n'importe où dans les 24 heures qui précèdent.

12. Donner la formule qui permet de simuler, sur les données de l'expérience, des jeux de données *supplémentaires* à chacune des trois températures pour tenir compte de cette réalité, en utilisant des simulations d'une loi uniforme $[0 ; 1]$. Des données gaussiennes plutôt qu'uniformes seraient-elles préférables à celles-ci ?

Dans la suite, on suppose que vous avez simulé 5 jeux de données supplémentaires pour chacune des 3 températures.

13. Obtenir, comme à la première partie, 15 nouvelles valeurs des paramètres $\beta, \theta, t_{0,05}$ aux diverses températures. Commenter sur la stabilité des estimateurs qu'on peut attendre de ces calculs, à l'aide de statistiques descriptives.
14. Obtenir du modèle d'Arrhenius, comme à la deuxième partie, à partir des 15 valeurs de $\ln(t_{0,05})$ trouvées à la question précédente, une estimation de $t_{0,05}$ à 50° , son intervalle de prédiction à droite de niveau 95%, ainsi que la proportion estimée des condensateurs tels $t_{0,05}$ dépasse 10 ans. Donner cette proportion en pourcentage. Considérez-vous que ce pourcentage soit acceptable? Donner les limites de cette estimation.
15. Admettons que les résultats précédents soient bien validés. La compagnie N*** vous semble-t-elle en mesure selon vous de respecter le cahier des charges de son principal contrat qui stipule que ses condensateurs catalytiques seront garantis dix ans? Vous procéderez par un test d'hypothèse.
16. Pour obvier jusqu'à un certain point au défaut de validité constaté plus haut, on peut penser utiliser, choisie au hasard, une observation simulée à chacune des trois températures (telle qu'obtenue à la section), parmi les 5 modèles de Weibull (obtenus comme à la section), pour effectuer une analyse de régression selon la loi d'Arrhenius. Donc, une régression sur 3 points, un à chaque température. De combien de valeurs extrapolées de $t_{0,05}$ pourrions-nous alors disposer, avec les 5 simulations par température?
17. Selon vous, serait-il possible d'utiliser ces données simulées de la probabilité de survie à 10 ans, pour en estimer le 5^e centile? Serait-il utile d'avoir plus de simulations? Verriez-vous un avantage à cette méthode sur celle utilisant les 15 simulations des données originales dans le modèle de régression d'Arrhenius comme à la question 14?
18. Utilisez 10 valeurs prédites calculées de $t_{0,05}$ à 50° calculées selon cette nouvelle procédure, pour trouver un intervalle de tolérance à droite de couverture 95% pour ce paramètre. Énoncez vos hypothèses et donnez-en les limites, dont la confiance à accorder à cet intervalle.

Exercices facultatifs.

19. Décrivez une technique de simulation qui permettrait de valider les résultats dans le cas où les moments des pannes sont déterminés avec précision et non seulement constatés comme ici. Pourquoi une étude de validation dans ce cas-là ?
20. La précision des estimateurs est évidemment fonction de l'information disponible. Eût-on décidé que l'expérience durerait seulement 6 mois ou 182 jours, pourrait-on croire que les estimateurs seraient aussi valables ? A-t-on assez d'information ici pour porter un jugement ?

Les techniques récentes de simulation et de rééchantillonnage dont on donne un petit aperçu ici sont devenues très pratiquées, voire essentielles, aujourd'hui dans le traitement des données, tout particulièrement pour les estimations.