Comparaison des lois Lognormales et Weibull généralisée dans le modèle AFT

Luc Clerjaud

Weibull Généralisé et Log-normale dans le modèle AFT.

Luc CLERJAUD, *Consultant.*

Résumé
Les travaux précédents ont montré que la loi de Weibull Généralisé présente une meilleure qualité d'ajustements de la fonction de survie que la loi Lognormale lorsque les taux d'hasard sont en forme de $\cup$ *, dans le modèle AFT. Or on a vu que ces lois ajustent presque de même qualité cette fonction de survie lorsque les taux d'hasard ont la forme en* $\cap$*. Les tests d'hypothèses seront envisagés pour montrer que la loi de Weibull Généralisée est largement meilleure en qualité régression lorsque la fonction de risque est en forme de* $\cup$*.*

Abstract
We have seen, according to these propositions, the Generalized Weibull distribution of life-time duration fit better these survival function than the Log-normal distribution, when the hazard rate function has a $\cup - shaped$*. AFT-Lognormal and AFT-Generalized Weibull fit better these survival functions in case of* $\cap$*-shaped hazard functions according to the last studies, but these models fits with the same quality of fits. We propose testing these hypothesis by using hypothesis tests in order to show how this GW distribution fits significanty better theses life-time durations than Log-normal distribution. Then, we will test Log-normal against WG distributions when the hazard rate have a* $\cup$*- shape.*

Les hypothèses de distributions des temps de survie sont suggérés à l'heure actuelle, tel que les lois exponentiels(fonction de risque constant), log-normal(fonction de risque monotone), et log-logistiques avec des variables constants dans le temps([BagNik02], [Law03], [ME98],[Nel04], [Vie88], [Sin71]).Ces modèles peuvent être aussi utilisés lorsque les courbes de taux d'hasard ont une forme en $\cup$. Nous considérons l'application de la loi de Weibull Généralisé (GW) [BagNik02], dans le modèle ALT. ([Bag78], [Nel04]).

On se propose de tester les hypothèses, soit les temps de survie suivent une loi Log-normale ou soit ils suivent une loi de Weibull Généralisé, pour montrer le degré de significativité de la qualité d'ajustement de la fonction de surtvie par rapport ` la loi Log-normale.

1.1. Estimation dans le modèle AFT-GW

1.1.1. *Le modèle AFT : définition*

Soit $x(\cdot) = (x_0(\cdot), \ldots, x_m(\cdot))^T$, $x^{(0)}(t) \equiv 1$, le vecteur des variables à m dimension, dynamiques dans le temps, où x_i est à une dimension, $i = 1, \ldots, m$, et $x^{(0)}$ le vecteur des stress ou covariantes usuels(température,...) à m dimensions (standard,normal) qui varient dans le temps.Supposons que la durée de vie $T_{x(\cdot)}$ sous stress $x(\cdot)$ est une variable aléatoire qui est positive, et $S_{x(\cdot)}$ la fonction de survie de $T_{x(\cdot)}$:

$$S_{x(\cdot)}(t) = P\left\{T_{x(\cdot)} > t\right\} \tag{1.1}$$

Le modèle AFT est vérifié dans l'ensemble E des covariantes(ou stress) s'il existe une fonction $r : E \longrightarrow R_+$ tel que

$$S_{x(\cdot)}(t) = S_0\left(\int_0^t r[x(\tau)]d\tau\right),\ \forall\, x(\cdot) \in E, \tag{1.2}$$

où S_0 est la fonction de survie correspondant à des conditions normales et standards.Si $x(\tau) = x \in E_1$ est constant dans le temps, alors

$$S_x(t) = S_0(r(x)t),\ \forall\, x \in E_1. \tag{1.3}$$

Dans le cas paramétrique, S_0 appartient à une *famille de fonctions paramétrisées*, et donc $r(x)$ est paramétrisé.

On considère les fonctions de survie sous stress usuelle S_0, suivant les lois distribution, telles que:

$$S_0(t) = 1 - \Phi\left(\nu \ln t\right),\ \nu > 0,\ \ (\textit{Lognormal}) \tag{1.4}$$

où $\Phi(u)$ est la fonction de répartition de la loi Normale,

$$S_0(t) = exp\left\{1 - (1 + t^{\nu})^{\gamma}\right\},\ \nu, \gamma > 0,\ \ (\textit{Generalized Weibull}). \tag{1.5}$$

En tenant compte des lois Lognormal et Weibull Généralisé en tant qu'hypothèse de loi de distribution dans le modèle AFT(1.2), nous obtenons les modèles AFT-lognormal et AFT-GW respectivement.Les estimations de la fonctions de survie sont en fonction des plans d'expériences : premiers plan d'expérience, deuxième plan d'expérience, etc...

1.1.2. *Estimation des paramètres et des fonctions de survie: modèles AFT-GW*

Si on considère le premier plan d'expérience, sous des stress constant : n_i unités sont tests dans des conditions où les covariantes sont $x^{(i)}$ $(i = 1, \ldots, k)$. Soit t_i la durée maximale d'expérimentation pour le i-ème échantillon. Soit $\beta = (\beta_0, \ldots, \beta_m)^T$ un vecteur de paramètre de régression. La durée de vie des j-èmes unités du i-ème groupe est noté T_{ij}. Soit $X_{ij} = T_{ij} \wedge t_i$ et $\delta_{ij} = \mathbf{1}\{T_{ij} < t_i\}$. La fonction de vraissemblance est:

$$\begin{aligned} L(\beta,\nu,\gamma) &= \prod_{i=1}^{k}\prod_{j=1}^{n_i}\left\{\nu\gamma e^{-\nu\beta^T x^{(i)}} X_{ij}^{\nu-1}\left(1+\left(e^{-\beta^T x^{(i)}} X_{ij}\right)^{\nu}\right)^{\gamma-1}\right\}^{\delta_{ij}} \times \\ &\times \quad exp\left\{1-\left(1+\left(e^{-\beta^T x^{(i)}} X_{ij}\right)^{\nu}\right)^{\gamma}\right\} \end{aligned} \tag{1.6}$$

où $x^{(i)} = (x_{i0}, \ldots, x_{im})$, $x_{i0} = 1$.

La fonction de scores est exprimée par $U_l(\beta,\nu,\gamma)$ et la matrice d'information de Fischer par $I(\beta,\nu,\gamma) = (I_{ls}(\beta,\nu,\gamma)))_{(m+3)\times(m+3)}$, où $I_{ls}(\beta,\nu,\gamma)$ est moins la dérivée partielle de $\ln L$.

L'estimateur des fonctions de survies sous stress usuelle $x^{(0)}$ peut être égale à :

$$\hat{S}_{x^{(0)}}(t) = exp\left\{1-\left(1+\left(e^{-\hat{\beta}^T x^{(0)}} t\right)^{\hat{\nu}}\right)^{\hat{\gamma}}\right\} \tag{1.7}$$

avec ses intervalles de confiances de niveau $(1-\alpha)$. Dans le cas du second plan d'expérience, l'estimation des paramètres, de la fonction de survie et ses intervalles de confiances de niveau $(1-\alpha)$ sont obtrnu par la maximisation de la fonction de vraissemblance qui, issue des propriétés de Sedyakin, est exprimée par:

$$\begin{aligned} L(\beta,\nu,\gamma) &= \prod_{i=1}^{n}\left\{\nu\gamma e^{-\nu\beta^T x^{(i)}(X_i)}(f_i(X_i,\beta,\gamma))^{\nu-1}\left(1+(f_i(X_i,\beta,\gamma))^{\nu}\right)^{\gamma-1}\right\}^{\delta_i} \times \\ & exp\left\{1-(1+(f_i(X_i,\beta,\gamma))^{\nu})^{\gamma}\right\}, \end{aligned} \tag{1.8}$$

où X_i est le moment de défaillance du i-ème unité et

$$f_i(t,\beta,\gamma) = \int_0^t e^{-\beta^T x^{(i)}(u)} du.$$

1.1.3. *Estimation des paramètres et des fonctions de survie: modèles AFT-LN*

L'estimation est identique que celle du modèle AFT-WG(pour Weibull Généralisé).La fonction de vraissemblance pour ce modèle AFT-Lognormal est défine, dans le cas du premier plan d'expérience, par:

$$L(\beta,\sigma) = \prod_{i=1}^{k}\prod_{j=1}^{n_i}\left[\frac{1}{\sigma X_{ij}} h\left(\frac{\ln X_{ij} - \beta^T x^{(i)}}{\sigma}\right)\right]^{\delta_{ij}} G\left(\frac{\ln X_{ij} - \beta^T x^{(i)}}{\sigma}\right) \tag{1.9}$$

oú $h(u) = -G'(u)$, $\lambda(u) = \frac{h(u)}{G(u)}$,$u \in R$. $G(u) = 1 - \Phi(u)$. $\Phi(u)$ est la fonction de répartition de la loi normale centrée réduite.

De même, grâce au calcul de scores et d'informations de Fischer, l'estimateur de la fonction de survie sous stress usuel $x^{(0)}$ peut être égale à:

$$\hat{S}_{x^{(0)}}(t) = exp\left\{1 - \left(1 + \left(e^{-\hat{\beta}^T x^{(0)}} t\right)^{\hat{\nu}}\right)^{\hat{\gamma}}\right\} \tag{1.10}$$

avec ses intervalles de confiances de niveau $(1 - \alpha)$. Après estimation, nous mettons

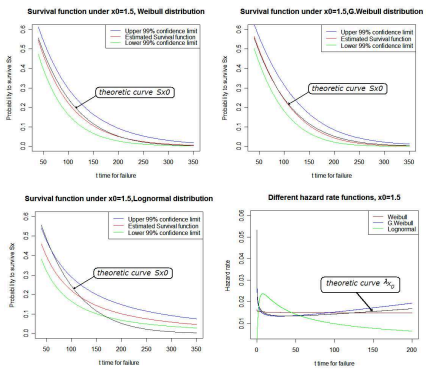

Figure 1.1. *Point and interval estimators of survival and hazard rate functions under usual stress $x^{(0)}$.*

en évidence l'importance d'écart d'ajustement des données de survie entre ces deux lois de distribution(Log-normal, WG) en terme d'ajustement de la fonction de survie lorsque les taux de défaillance sont en forme de $\cup$. Pour cela, nous allons utiliser le test du Khi-deux avec les classes de Neyman-Pearson[Pr96].

1.2. Loi Lognormal ou WG : Test d'hypothèses

Supposons 2 hypothèses de distributions des durées de vies lorsque le taux d'hasard est en forme de $\cup$. On suppose qu'il y a r classes dans un échantillon de taille n (par exemple n durées de vie enregistrées). Soit un vecteur de fréquences ν =$(\nu_1, ...\nu_r)^T$ et p = $(p_1, ..., p_r)^T$ un vecteur de probabilités basé sur les hypothèses des lois de distributions. Mais si les varaibles alèatoires T_0,T_1,...,T_n correspondant au durèes de vie,sont i.i.d, alors :

$$P\left(T_i < t\right) = F(x) \tag{1.11}$$

où F(x) est inconnue.

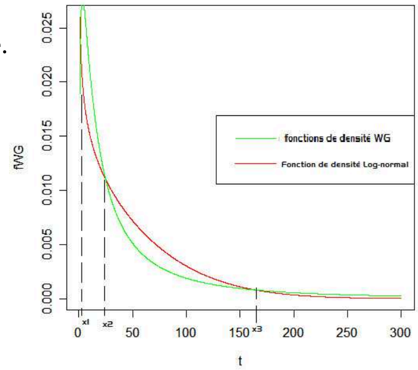

Figure 1.2. *Intercept points between two density functions :*

Dans le modèle AFT, on suppose 2 hypothèses. H_0 est l'hypothèse où les durée de vie suivent une loi Log-normale et H_1 est l'hypothèse où elles suivent une loi WG. C'est à dire : $H_0 : f(t) = f_{LN}(t)$ contre $H_1 : f(t) = f_{WG}(t)$, où $f_{WG}(t)$ est la fonction de densité de probabilité de la loi de Weibull Généralisée et $f_{LN}(t)$ est la fonction de densité de probabilité de la loi Lognormale.

Les classes de Neymann-Pearson sont déterminées par les bornes $\alpha_0, \alpha_1, ..., \alpha_k$ où α_j est le point d'intersection d'ordre j vérifiant: $f_{WG}(t) = f_{LN}(t)$. Soient I_1 et I_2, 2 classes telles que : $I_1 = \{t : f_{WG} \geq f_{LN}\}$ de cardinal ν_1 et $I_2 = \{t : f_{WG} < f_{LN}\}$ de cardinal ν_2. La statistique utilisée pour 2 classes seulement est la statistique de Rao-Robson-Nikulin:

$$Y_n^2(\widehat{\theta}) = X_n^2(\widehat{\theta}) + \frac{1}{n}\Lambda^T(\widehat{\theta})\left(i(\widehat{\theta}) - J(\widehat{\theta})\right)^{-1}\Lambda(\widehat{\theta}) \tag{1.12}$$

où

$$\beta(\theta) = \begin{pmatrix} \beta_{11}(\theta) \dots \beta_{1m}(\theta) \\ \vdots \qquad\quad \vdots \\ \beta_{r1}(\theta) \dots \beta_{rm}(\theta) \end{pmatrix}$$

m est la dimension de θ et :

$$\beta_{ij}(\theta) = \frac{1}{\sqrt{p_i(\theta)}}\frac{\partial p_i}{\partial \theta_j},$$

$\Lambda(\theta) = (\Lambda_1(\theta), ..., \Lambda_m(\theta))^T$ avec :

$$\Lambda_j(\theta) = \sum_{i=0}^{r} \frac{\nu_i \partial p_i}{p_i \partial \theta_j},$$

pour $i = 1, ..., r$ et $j = 1, ..., s$. Où $p_i(\theta) = \int_{a_{i-1}}^{a_i} f(t, \theta) dt$, $i = 1, ..., r$. $ni(\theta)$ est la matrice d’information de Fischer de l’échantillon and $J(\theta) = \beta^T(\theta)\beta(\theta)$ est la matrice d’information de Fischer des données groupées. Sous H_0 les temps de survies suivent la loi de Weibull Généralisée, et la statistiques Y_n^2 suit asymptotiquement la loi χ_1^2. On rejette H_0 avec un niveau de risque $\alpha = 0,01$ si $Y_n^2 > c_\alpha$

On obtient alors : $\nu_1 = 220$ et $\nu_2 = 130$ et $p_1 = 0,6218152$ et $p_2 = 0,3781848$. Après estimation de la statistique de Rao-Robson-Nikulin, on obtient $Y_n^2 = 0,3616$ et la signification est de $0,5475672 > 0,01$. On ne peut rejeter l’hypothèse H_0. Donc la loi WG ajuste largement mieux que la loi Log-normale lorque les taux d’hasard ont la forme en $\cup$.

1.3. conclusion

Les tests d’hypothèses ont largement péblicité la loi WG étant la loi de distribution des duré de vie qui ajuste le mieux lorsque les taux d’hasard ont la forma en $\cup$. On le retrouve dans les travaux précédents. Ce n’est pas surprenant que la loi WG ajuste largment mieux la fonction de survie que la loi Log-normale. On le retrouve ce résultat sur le graphique (voir figure). Ultérieurement, on ira tester la loi WG contre a loi Log-normale par rapport à l’ajustement de la loi de suruvie dans le modèle AFT, lorsque les taux d’hasard ont la forme en $\cap$, ultérieurement.

1.4. Bibliography

[Bag78] Bagdonavičius, V. (1978) Testing the hyphothesis of the additive accumulation of damages. *Probab. Theory and its Appl.*, **23**, No. 2, 403-408.

[Nel04] Nelson, W. (2004). *Accelerated Testing, Statistical Models, Test Plans, and Data Analysis*. John Wiley & Sons, New Jersey. 601p.

[BagNik02] Bagdonavicius, V. and Nikulin, M. (2002) *Accelerated Life Models*. Chapman and Hall/CRC, Boca Raton.

[Law03] Lawless, J.F. (2003) *Statistical Models and Methods for Lifetime Data*. New York: John Wiley and Sons.

[ME98] Meeker W.Q, Escobar.L. (1998). *Statistical Method for Reliability Data*. J.Wiley, New York.

[Sin71] Singpurwalla, N.D. (1971) Inference from Accelerated Life Tests When Observations Are Obtained from Censored Samples, *Technometrics*, **13**, 161-170.

[Vie88] Viertl, R. (1988). Statistical Methods in Accelerated Life Testing. Vandenhoeck & Ruprecht, Gottingen.

[Pr96] Priscilla E.Greenwood, Mikhail S.Nikulin (1996) : A Guide to Chi-Squared Testing, Chapter 2.Wiley Series in Probability and Statistics. New York, Chichester, Bisbane, Toronto, Singapore.