

TP N° 41

Estimation d'un intervalle de confiance par inversion de la matrice de Fisher

L'objet de ce TP¹ est d'estimer des intervalles de confiance par inversion de la matrice de Fisher à l'issue d'un ajustement par la méthode du maximum de vraisemblance.

Portant sur les paramètres d'une loi de Weibull et sur un quantile de la population, les résultats obtenus par dérivation de l'expression analytique du logarithme de la fonction de vraisemblance sont comparés à ceux fournis par l'outil GEN CAB qui utilise une méthode exclusivement numérique.

- 1 – Présenter la méthode d'estimation des intervalles de confiance par la matrice de Fisher
- 2 – Appliquer cette méthode à une la loi de Weibull à 3 paramètres en dérivant l'expression de la log vraisemblance.
- 3 – A partir d'un échantillon simulé, estimer les intervalles de confiance à 90% encadrant les paramètres de la loi ainsi que le quantile à 99 % de la population.
- 4 – Proposer une méthode de calcul discrète.
- 5 – Comparer les résultats obtenus par les deux méthodes de traitement.

¹ Ce TP a été élaboré en collaboration avec Anaïs Tanguy, une brillante stagiaire.

1 - Estimation des intervalles de confiance par la matrice de Fisher

L'information de Fisher quantifie l'information relative à un paramètre contenue dans une distribution.

La technique d'estimation par le maximum de vraisemblance consiste à choisir la valeur du paramètre qui maximise la vraisemblance des observations.

L'information de Fisher est définie comme la variance associée au maximum du logarithme de la vraisemblance.

$$I_n(\theta) = E \left\{ \left(\frac{\partial \ln L(X, \theta)}{\partial \theta} \right)^2 \right\}$$

Si la log vraisemblance est deux fois dérivable et sous certaines conditions de régularité, l'information de Fisher peut aussi s'écrire :

$$I_n(\theta) = -E \left(\frac{\partial^2 \ln L(X, \theta)}{\partial \theta^2} \right)$$

Dans le cas d'une loi à plusieurs paramètres, on ne parle plus d'information de Fisher mais de matrice de Fisher :

$$F_{ij} = I_n(\theta)_{ij} = E \left\{ \left(\frac{\partial \ln L(X, \theta)}{\partial \theta_i} \right) \left(\frac{\partial \ln L(X, \theta)}{\partial \theta_j} \right) \right\} = -E \left(\frac{\partial^2 \ln L(X, \theta)}{\partial \theta_i \partial \theta_j} \right)$$

Soit dans le cas particulier à 3 paramètres :

$$F = I_n(\theta) = \begin{pmatrix} -\frac{\partial^2 \ln L(X, \theta)}{\partial \theta_1^2} & -\frac{\partial^2 \ln L(X, \theta)}{\partial \theta_1 \partial \theta_2} & -\frac{\partial^2 \ln L(X, \theta)}{\partial \theta_1 \partial \theta_3} \\ -\frac{\partial^2 \ln L(X, \theta)}{\partial \theta_2 \partial \theta_1} & -\frac{\partial^2 \ln L(X, \theta)}{\partial \theta_2^2} & -\frac{\partial^2 \ln L(X, \theta)}{\partial \theta_2 \partial \theta_3} \\ -\frac{\partial^2 \ln L(X, \theta)}{\partial \theta_3 \partial \theta_1} & -\frac{\partial^2 \ln L(X, \theta)}{\partial \theta_3 \partial \theta_2} & -\frac{\partial^2 \ln L(X, \theta)}{\partial \theta_3^2} \end{pmatrix}$$

L'inverse de la matrice de Fisher est la matrice de variance-covariance.

La racine carrée des éléments de la diagonale correspond à l'écart-type de chacun des paramètres et permet de définir un intervalle de confiance encadrant celui-ci en considérant une distribution normale.

De même pour une fonction g des différents paramètres, correspondant par exemple à la valeur d'un quantile, la variance s'exprime de la manière suivante : $\sigma_{\hat{g}_n}^2(\theta) = \nabla g(\theta)^T I_n^{-1}(\theta) \nabla g(\theta)$

Avec $\nabla g(\theta)$ et $\nabla g(\theta)^T = \left(\frac{\partial g(\theta)}{\partial \theta_1} ; \dots ; \frac{\partial g(\theta)}{\partial \theta_m} \right)$ le gradient de la fonction g et sa transposée

Et $I_n^{-1}(\theta)$ l'inverse de la matrice de Fisher.

2 - Application à la loi de Weibull

La densité de probabilité de la loi de Weibull a pour expression :

$$f(X_i; \beta, \sigma, \gamma) = \frac{\beta}{\sigma} \left(\frac{X_i - \gamma}{\sigma}\right)^{\beta-1} \exp\left(-\left(\frac{X_i - \gamma}{\sigma}\right)^\beta\right)$$

Dans le cas de données non censurées, la vraisemblance est égale au produit de la densité de probabilité des observations :

$$L(X_1, \dots, X_n; \beta, \sigma, \gamma) = \prod_{i=1}^n f(X_i; \beta, \sigma, \gamma) = \left(\frac{\beta}{\sigma}\right)^n \left[\prod_{i=1}^n \left(\frac{X_i - \gamma}{\sigma}\right)^{\beta-1} \right] \exp\left(-\sum_{i=1}^n \left(\frac{X_i - \gamma}{\sigma}\right)^\beta\right)$$

La log vraisemblance a pour expression :

$$\ln L(X_1, \dots, X_n; \beta, \sigma, \gamma) = n \ln(\beta) - n \ln(\sigma) + \sum_{i=1}^n \left[(\beta - 1) \times \ln(X_i - \gamma) - (\beta - 1) \times \ln(\sigma) - \left(\frac{X_i - \gamma}{\sigma}\right)^\beta \right]$$

La double dérivation de cette expression par rapport aux paramètres β , σ et γ conduit à la matrice de Fisher suivante, sous une forme analytique :

$$I_n(\beta, \sigma, \gamma) = \begin{pmatrix} \frac{n}{\beta^2} + \sum_{i=1}^n \left(\frac{X_i - \gamma}{\sigma}\right)^\beta [Ln\left(\frac{X_i - \gamma}{\sigma}\right)]^2 & \frac{n}{\sigma} - \sum_{i=1}^n \left[\frac{(X_i - \gamma)^\beta}{\sigma^{\beta+1}} \times (1 + \beta \times Ln\left(\frac{X_i - \gamma}{\sigma}\right)) \right] & - \sum_{i=1}^n \frac{(X_i - \gamma)^{\beta-1}}{\sigma^\beta} \left[1 + \beta \times Ln\left(\frac{X_i - \gamma}{\sigma}\right) \right] - \frac{1}{(X_i - \gamma)} \\ \frac{n}{\sigma} - \sum_{i=1}^n \left[\frac{(X_i - \gamma)^\beta}{\sigma^{\beta+1}} \times (1 + \beta \times Ln\left(\frac{X_i - \gamma}{\sigma}\right)) \right] & \frac{-n\beta}{\sigma^2} + \frac{\beta(\beta+1) \sum_{i=1}^n (x_i - \gamma)^\beta}{\sigma^{\beta+2}} & \frac{\beta^2}{\sigma^{\beta+1}} \sum_{i=1}^n (X_i - \gamma)^{\beta-1} \\ - \sum_{i=1}^n \frac{(X_i - \gamma)^{\beta-1}}{\sigma^\beta} \left[1 + \beta \times Ln\left(\frac{X_i - \gamma}{\sigma}\right) \right] - \frac{1}{(X_i - \gamma)} & \frac{\beta^2}{\sigma^{\beta+1}} \sum_{i=1}^n (X_i - \gamma)^{\beta-1} & -(1 - \beta) \sum_{i=1}^n \frac{1}{(X_i - \gamma)^2} - \frac{\beta(\beta-1)}{\sigma^\beta} \sum_{i=1}^n (X_i - \gamma)^{\beta-2} \end{pmatrix}$$

Après remplacement dans celle-ci de la valeur des paramètres obtenus à l'issue de l'ajustement par le maximum de vraisemblance et de celle des diverses observations, cette matrice peut être inversée numériquement pour obtenir la matrice de variance-covariance.

L'expression analytique d'un quantile peut se calculer en inversant la fonction de répartition de la loi de Weibull $\exp\left(-\left(\frac{X_i - \gamma}{\sigma}\right)^\beta\right)$ soit : $Q(p; \beta, \sigma, \gamma) = \lambda - \sigma \times Ln(p)^{1/\beta}$

3 - Méthode de calcul discrète

La méthode de calcul utilisée par GENCAB est une méthode discrète fondée sur les formules de Taylor². Elle permet de traiter indifféremment des lois de probabilité diverses (Bertholon, multi-Weibull, loi généralisée des extrêmes, etc.) qui peuvent être éventuellement couplées à des lois d'accélération (Arrhenius, Eyring, Cox, etc.).

Le calcul des dérivées secondes de la diagonale de la matrice de Fisher s'effectue de la manière suivante :

$$\frac{\partial^2 f}{\partial \theta_i^2}(\hat{\theta}_i) = \lim_{h \rightarrow 0} \frac{f(\hat{\theta}_i + h) + f(\hat{\theta}_i - h) - 2f(\hat{\theta}_i)}{h^2}$$

² Cette méthode a fait l'objet d'une proposition de communication au congrès ESREL 2012.

qui résulte de la sommation des expressions suivantes :

$$f(\hat{\theta}_i + h) = f(\hat{\theta}_i) + \frac{\partial f}{\partial \theta_i}(\hat{\theta}_i) \times h + \frac{\partial^2 f}{\partial \theta_i^2}(\hat{\theta}_i) \times \frac{h^2}{2} + \varepsilon(h)$$

$$f(\hat{\theta}_i - h) = f(\hat{\theta}_i) - \frac{\partial f}{\partial \theta_i}(\hat{\theta}_i) \times h + \frac{\partial^2 f}{\partial \theta_i^2}(\hat{\theta}_i) \times \frac{h^2}{2} + \varepsilon(h)$$

De même, le calcul d'un quantile s'effectue de manière numérique, itérative, par inversion de la fonction de répartition du modèle probabiliste.

4 - Comparaison des résultats avec ceux fournis par l'outil

L'ajustement réalisé ci-dessous par l'outil GEN CAB donne directement les intervalles de confiance souhaités.

Le calcul analytique de la matrice de Fisher est également réalisé sur la feuille de calcul (non visible à l'écran). L'erreur relative des termes de la matrice est inférieure à $3 \cdot 10^{-3}$.

Ajustement *Maximum de vraisemblance*

Loi de probabilité : **WEIBULL (3 paramètres)**

Bêta : 2,70389632
 Sigma : 7,6776401
 Gamma : -1,52235458

LN Vraisemblance

-121,370574

Taux de confiance : 90%

Non censurées

LN K (non censurées)

-121,370574

Détail des calculs analytiques

↓

EEEEEB
 dccccd

	Min	Max	
Bêta :	2,703896317	0,50528838	4,902504254
Sigma :	7,677640101	2,703751361	12,65152884
Gamma :	-1,522354577	-6,02620989	2,981500736

	Min	Max	
Quantile 99 :	11,9834457	9,528835356	14,43805604

	Min	Max	
Matrice de Fisher :	11,05466305	-2,546190327	2,459007684
	-2,546190327	6,201498051	5,573979975
	2,459007684	5,573979975	7,395179888

	Min	Max	
Matrice de variance-covariance :	1,786656852	3,929887156	-3,55616978
	3,929887156	9,144031198	-8,19889041
	-3,55616978	-8,198890415	7,497464652

Matrice de Fisher par calcul analytique

11,05464779	-2,546111519	2,457993358
-2,546111519	6,201468908	5,573579577
2,457993358	5,573579577	7,393013557

Erreur absolue

1,52532E-05	-7,88078E-05	0,001014325
-7,88078E-05	2,91433E-05	0,000400398
0,001014325	0,000400398	0,002166332

Erreur relative

1,3798E-06	3,09522E-05	0,000412664
3,09522E-05	4,69942E-06	7,18386E-05
0,000412664	7,18386E-05	0,000293024

Variable	Taux : λ(t)	R(t) = 1-F(t)	Densité : f(t)	Ln(f(t))
6,52595539	0,38163911	0,32110993	0,12254811	-2,09925161
9,52456046	0,65463693	0,06893752	0,04512905	-3,09822918
5,7520005	0,32124357	0,42136786	0,13536172	-1,9998047
9,63908222	0,66624262	0,06391581	0,04258343	-3,15628999
3,70601041	0,18299827	0,70197817	0,12846079	-2,05213155
1,09229389	0,05618948	0,94711494	0,05321789	-2,93336063
1,70024967	0,08023391	0,90880406	0,07291691	-2,61843476
6,93861722	0,41557932	0,27241695	0,11321085	-2,17850328
6,34219357	0,36691142	0,343971	0,12620689	-2,06983274
8,05106754	0,51293553	0,16265875	0,08343345	-2,48370595
8,03959799	0,51188888	0,16361753	0,083754	-2,47987141
9,28455031	0,63058824	0,08043317	0,05072021	-2,98143076
4,60432932	0,23975991	0,58084848	0,13926418	-1,97138258
6,59109407	0,38691705	0,31317192	0,12117156	-2,11054791
3,77534805	0,18715272	0,6930276	0,129702	-2,04251578
2,87705661	0,13636521	0,80101559	0,10923066	-2,21429348
7,72169119	0,48323105	0,19165391	0,09261312	-2,37932444
0,70754065	0,04284169	0,96528545	0,04135446	-3,18557501
1,65377463	0,07827236	0,91215755	0,07139672	-2,63950333
8,38981798	0,54424497	0,1359947	0,07401443	-2,60349521
0,67548422	0,04179761	0,96659584	0,04040139	-3,208891
5,9036387	0,33273727	0,40098503	0,13342267	-2,01423325
8,94038464	0,59675494	0,09934606	0,05928525	-2,82539472
4,31530591	0,22080937	0,62081433	0,13708162	-1,98717873
8,20270739	0,52685628	0,15032798	0,07920124	-2,5357633
3,55926869	0,17433355	0,7206243	0,12562899	-2,0744222
0,47000398	0,0353604	0,9742813	0,03445098	-3,36821796



Fisher-Weibull

La feuille de calcul Excel est accessible par un clic de souris sur l'icône :