Sur les applications de la famille des lois Gaussiennes inverses en fiabilité.

Noureddine Saadia

Sur les applications de la famille des lois Gaussiennes inverses en fiabilité

N. Saaidia
Université Victor Segalen Bordeaux 2, France &
Université Badji Mokhtar, Annaba, algérie

1 Introduction

La distribution Gaussienne inverse (IGD) tire son nom du fait qu'elle établit une relation avec la distribution normale. L'IGD est introduit la première fois par Schrodinger en 1915 [Seshadri 1993]. Il est bien reconnu que l'IGD a plusieurs applications intéressantes en biologie, économie, cardiologie, hydrologie, démographie, finance, linguistique, etc. Pour plus de détails sur l'IGD, voir par exemple [Seshadri 1993], [Seshadri 1999], [Chhikara and Folks, 1989], [Voinov and Nikulin 1993]. L'IGD s'est avérée très concurrente à la distribution de Weibull, Weibull généralisée, ainsi qu'à la distribution log-normale.
Nous étudions des possibilités (perspectives) d'applications de l'IGD en fiabilité, et en particulier, la validations des modèles basés sur l'IGD.

2 Propriétés et Caractérisation

Soit $\mathbb{X} = (X_1, X_2, \cdots, X_n)^T$ un n-échantillon, i.e: n varibles aléatoires indépendantes et identiquement distributées (iid), on dit que X_i suit la distribution Gaussienne inverse (et on note $X_i \sim IG(\mu, \lambda)$) si sa fonction de densité est définie par:

$$f(x, \mu, \lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} exp\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\}, \quad x \geq 0, \quad \mu > 0, \quad \lambda > 0.$$

où μ est la moyenne et λ est le paramètre de forme.
Cette fonction est unimodale. On peut démonter que

$$E(X_i) = \mu \ , Var(X_i) = \frac{\mu^3}{\lambda}$$

3 Estimation des paramètres

La fonction de vraisemblance $L(\mathbb{X}, \theta)$, avec $\theta = (\mu, \lambda)^T$ de l'échantillon $\mathbb{X} = (X_1, X_2, \cdots, X_n)^T$ est donnée par la formule suivante:

$$L(\mathbb{X}, \theta) = \prod_{i=1}^{n} f(X_i, \theta) = \lambda^{\frac{n}{2}} (2\pi)^{\frac{n}{2}} \prod_{i=1}^{n} X_i^{-\frac{3}{2}} exp\left(\sum_{i=1}^{n} \frac{-\lambda(X_i - \mu)^2}{2\mu^2 X_i}\right)$$

Posons: $V = \sum_{i=1}^{n}(X_i^{-1} - \overline{X}^{-1})$ donc, Les estimateurs du maximum de vraisemblance (MV) de μ et λ sont:

$$\begin{cases} \hat{\mu} = \overline{X} \\ \hat{\lambda} = \dfrac{n}{\sum_{i=1}^{n}(X_i^{-1} - \overline{X}^{-1})} = \dfrac{n}{V} \end{cases} \tag{1}$$

Les estimateurs sans biais de variance minimale de μ et λ sont respectivement:

$$\begin{cases} \hat{\mu} = \overline{X} \\ \hat{\lambda} = \dfrac{n-1}{V} \end{cases}$$

Notons que pour la famille $IG(\mu, \lambda)$, les statistiques $\overline{X}$ et V sont indépendantes, et que $T = \left(\overline{X}, V\right)^T$ est une statistique bidimensionnelle exhaustive minimale et complète, voir par exemple [Voinov and Nikulin 1993], [Seshadri 1993].

4 Tests d'adéquation pour l'IGD

Considérons le problème de tester l'hypothèse H_0 selon laquelle la distribution de $\mathbb{X} = (X_1, X_2, \cdots, X_n)^T$ vérifie:

$$H_0 : P(X_i \leq x) = F(x, \theta).$$

où $F(x, \theta)$ est la fonction de distribution de l'IGD.
Divisons la droite réelle en r intervalles : $I_1, I_2, \cdots, I_r$ mutuellement disjoints, par les poits:

$$0 = a_0 < a_1 < \cdots < a_{r-1} < a_r = +\infty.$$

Groupons l'échantillon $\mathbb{X} = (X_1, X_2, \cdots, X_n)^T$, nous obtenons le vecteur de fréquences $\nu = (\nu_1, \nu_2, \cdots, \nu_r)^T$.

Supposons que: $p_i = \int_{a_{i-1}}^{a_i} f(x, \theta)dx, \ i = 1, 2, \cdots, r$

et désignons par $X_n(\theta)$, le vecteur de composantes $\dfrac{\nu_i - np_i}{\sqrt{np_i}}, i = 1, 2, \cdots, r$.
L'information de Fisher de l'échantillon $\mathbb{X}$ est:

$$I_n(\theta) = ni(\theta) = \begin{pmatrix} \frac{n\lambda}{\mu^3} & 0 \\ 0 & \frac{n}{2\lambda} \end{pmatrix}.$$

Considérons la statistique de Rao-Robson-Nikulin ([Rao-Robson 1974], [Nikulin 1973], [Drost 1988])

$$Y_n^2(\hat{\theta}_n) = X_n^2(\hat{\theta}_n) + \frac{1}{n} l^T(\hat{\theta}_n)(i(\hat{\theta}_n) - J(\hat{\theta}_n))^{-1} l(\hat{\theta}_n).$$

Avec

$J(\theta) = B(\theta)^T B(\theta)$ est l'informationde de Fisher du vecteur de fréquences ν, où

$$B(\theta) = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \vdots & \vdots \\ b_{r1} & b_{r2} \end{pmatrix}; \ b_{i1}(\theta) = \frac{1}{\sqrt{p}_i} \frac{\partial p_i}{\partial \mu} \ , \ b_{i2}(\theta) = \frac{1}{\sqrt{p}_i} \frac{\partial p_i}{\partial \lambda}, \ \ i = 1, 2, \cdots, r$$

$$l(\theta) = (l_1(\theta), l_2(\theta))^T; \ \ l_1(\theta) = \sum_{i=1}^{r} \frac{\nu_i}{p_i} \frac{\partial p_i}{\partial \mu}(\theta), \ \ l_2(\theta) = \sum_{i=1}^{r} \frac{\nu_i}{p_i} \frac{\partial p_i}{\partial \lambda}(\theta).$$

Sous H_0, et pour n suffissament grand, la statistique Y_n^2 suit la distribution χ_{r-1}^2 à $(r-1)$ degrés de liberté [Greenwood and Nikulin 1996].

Cas 1. $\theta = (\mu, \lambda)^T$ est connu.

Dans ce cas la statistique

$$Y_n^2 = X_n(\theta)^T X_n(\theta) = X_n^2(\theta) = \sum_{i=1}^{r} \frac{(\nu_i - np_i)^2}{np_i}.$$

suit pour n assez grand, la distribution chi-deux χ^2_{r-1} à $(r-1)$ degrés de liberté.

Au seuil α, si on choisit la valeur critique

$$C_\alpha = \chi^2_{r-1,1-\alpha}.$$

L'hypothèse H_0 est acceptée si $Y_n^2 \le C_\alpha$. Dans le cas contraire, H_0 est rejetée.

Cas 2. μ ou λ est inconnu

On a alors

$$Y_n^2 = \begin{cases} X_n^2(\hat{\lambda}_n) + \frac{1}{n} \dfrac{\left(\sum_{i=1}^{r} \frac{\nu_i}{p_i} \frac{\partial p_i}{\partial \lambda}(\hat{\lambda}_n)\right)^2}{\frac{1}{2\hat{\lambda}_n} - \sum_{i=1}^{r} \frac{1}{p_i} \left(\frac{\partial p_i}{\partial \lambda}(\hat{\lambda}_n)\right)^2}, & \text{si } \mu \text{ est connu et } \lambda \text{ est inconnu} \\ X_n^2(\hat{\mu}_n) + \frac{1}{n} \dfrac{\left(\sum_{i=1}^{r} \frac{\nu_i}{p_i} \frac{\partial p_i}{\partial \mu}(\hat{\mu}_n))\right)^2}{\frac{\lambda}{\hat{\mu}^3} - \sum_{i=1}^{r} \frac{1}{p_i} (\frac{\partial p_i}{\partial \mu}(\hat{\mu}_n))^2}, & \text{si } \mu \text{ est inconnu et } \lambda \text{ est connu} \end{cases}$$

Au seuil α, si on choisit la valeur critique

$$C_\alpha = \chi^2_{r-1,1-\alpha}.$$

L'hypothèse H_0 est acceptée si $Y_n^2 \le C_\alpha$. Dans le cas contraire, H_0 est rejetée.

Cas 3. θ est inconnu.

L'estimateur de MV de $\widehat{\theta} = (\widehat{\mu}, \widehat{\lambda})^T$ est donné par (1). La statistique Y_n^2 peut s'écrire:

$$\begin{aligned} Y_n^2(\hat{\theta}_n) = X_n^2(\hat{\theta}_n) + \frac{1}{n \mid M \mid} & \left[\left(\frac{1}{2\hat{\lambda}_n^2} - \sum_{i=1}^{r} \frac{1}{p_i} \omega_{i2}^2(\hat{\theta}_n) \right) \left(\sum_{i=1}^{r} \frac{\nu_i}{p_i} \omega_{i1}(\hat{\theta}_n) \right)^2 \right] + \\ 2 & \left[\left(\sum_{i=1}^{r} \frac{1}{p_i} \omega_{i1}(\hat{\theta}_n) \omega_{i2}(\hat{\theta}_n) \right) \left(\sum_{i=1}^{r} \frac{\nu_i}{p_i} \omega_{i1}(\hat{\theta}_n) \right) \left(\sum_{i=1}^{r} \frac{\nu_i}{p_i} \omega_{i2}(\hat{\theta}_n) \right) \right] + \\ & \left[\left(\frac{\widehat{\lambda}_n}{\widehat{\mu}_n^3} - \sum_{i=1}^{r} \frac{1}{p_i} \omega_{i1}^2(\hat{\theta}_n) \right) \left(\sum_{i=1}^{r} \frac{\nu_i}{p_i} \omega_{i2}(\hat{\theta}_n) \right)^2 \right]. \end{aligned}$$

où

$$M = i(\hat{\theta}_n) - J(\hat{\theta}_n) = \begin{pmatrix} \frac{\hat{\lambda}_n}{\hat{\mu}_n^3} - \sum_{i=1}^{r} b_{i1}^2 & -\sum_{i=1}^{r} b_{i1} b_{i2} \\ -\sum_{i=1}^{r} b_{i1} b_{i2} & \frac{1}{2\hat{\lambda}_n^2} - \sum_{i=1}^{r} b_{i2}^2 \end{pmatrix}$$

$$\omega_{i1}(\hat{\theta}_n) = \frac{\partial p_i}{\partial \mu}(\hat{\theta}_n) \qquad et \qquad \omega_{i1}(\hat{\theta}_n) = \frac{\partial p_i}{\partial \lambda}(\hat{\theta}_n)$$

et $\mid M \mid$ est le déterminant de la matrice M.

Au seuil α, si on choisit la valeur critique

$$C_\alpha = \chi^2_{r-1,1-\alpha}.$$

L'hypothèse H_0 est acceptée si $Y_n^2 \le C_\alpha$. Dans le cas contraire, H_0 est rejetée.

5 Tests d'adéquation pour les données censurées

En fiabilité et en analyse de survie, nous confrontons souvent des observations incomplètes, et dans cette situation les méthodes habituelles ne sont plus valables. Dans le cas de censure aléatoire, plusieurs tests ont été proposés dont nous citons le test de Pearson modifié proposé par [Habib et Thomas (1996)]. Soit $(\mathbb{X}, \Delta) = (X_1, \delta_1), (X_2, \delta_2) \cdots, (X_n, \delta_n)$ un n-échantillon, tel que $X_i = min(T_i, C_i)$, T_i est le temps de panne (ou de survie), C_i est la censure et δ_i est la fonction indicatrice définie comme suit:

$$\delta_i = \begin{cases} 1 & si \quad T_i \leq C_i \\ 0 & sinon \end{cases}$$

Nous voulons tester l'hypothèse H_0 selon laquelle la distribution de $(\mathbb{X}, \Delta)$ vérifie:

$$H_0 : P(X_i > x) = S(x, \theta).$$

Où $S(x, \theta) = 1 - F(x, \theta)$ est la fonction de survie (ou de fiabilité) de l'IGD

Habib et Thomas (1996) ont démontré que $\sqrt{n}\left(\hat{S}_n(x) - S(x, \hat{\theta}_n)\right)$ converge vers un processus Gaussien, où $\hat{S}_n(x)$ est un estimateur de Kaplan-Meier et $\hat{\theta}_n$ est l'estimateur du maximum de vraisemblance.

Considérons maintenant une généralisation de la statistique de Pearson pour les données censurées aléatoires. En Divisons la droite réelle en r intervalles : $I_1, I_2, \cdots, I_r$ mutuellement disjoints, par les poits:

$$0 = a_0 < a_1 < \cdots < a_{r-1} < a_r = +\infty.$$

Soit le vecteur

$$\hat{Z}_n = \sqrt{n}\left(\hat{S}_n - S_{\hat{\theta}_n}\right)$$

Où $\hat{S}_n = (\hat{S}_n(a_1), \hat{S}_n(a_2), \cdots, \hat{S}_n(a_{r-1}))^T$ et $S_{\hat{\theta}_n} = (S(a_1, \hat{\theta}_1), S(a_2, \hat{\theta}_2), \cdots, S(a_{r-1}, \hat{\theta}_n))^T$.

La statistique

$$\hat{Q}_n = \hat{Z}_n^T \hat{\Sigma}^{-1} \hat{Z}_n.$$

Où $\hat{\Sigma}$ est la matrice des covariances, voir [Habib and Thomas 1996], suit pour n assez grand la distribution chi-deux χ^2_{r-1} à $(r-1)$ degrés de liberté.

Au seuil α, si on choisit la valeur critique

$$C_\alpha = \chi^2_{r-1, 1-\alpha}.$$

L'hypothèse H_0 est acceptée si $\hat{Q}_n \leq C_\alpha$. Dans le cas contraire, H_0 est rejetée.

Notons qu'à l'absence de la censure, la statistique $\hat{Q}_n$ se réduit à celle de Rao-Robson-Nikulin Y_n^2

6 La famille IGD et les modèles AFT

Dans cette section, nous considérons deux modèles de *survie* sur l'ensemble des stress E, $\{S_{x(\cdot)}, x(\cdot) \in E\}$, où $S_{x(\cdot)}$ est la fonction de *survie* de la variables aléatoire $T = T_{x(\cdot)}$ exprimant le temps de panne sous $x(\cdot) = (x_1(\cdot), ..., x_m(\cdot))^T$.

Rappelons que la fonction de *survie* et le taux de *hasard* sous $x(\cdot)$ sont :

$$S_{x(\cdot)}(t) = \mathbf{P}\left(T_{x(\cdot)} \geq t \mid x(s); 0 \leq s \leq t\right), \quad \lambda_{x(\cdot)}(t) = -\frac{S'_{x(\cdot)}(t)}{S_{x(\cdot)}(t)},$$

Nous avons le modèle de *fragilité* simple (*simple frailty model*) si le taux de hasard dépend non seulement de $x(\cdot)$, mais aussi d'une variable alétoire positive *non-observable* Z, appelée la *variable de fragilité*, de telle manière :

$$\lambda_{x(\cdot)}(t|Z=z) = z\lambda_0(t)r\{x(t)\}, \quad x(\cdot) \in E,$$

où $\lambda_0(\cdot)$ est le taux de hasard de base. Si nous supposons que la densité de la variable de *fragilité* Z appartient à la famille des lois Gaussiennes inverses $IG(\mu, \lambda)$, alors nous obtenons ce que l'on apppelle *le modèle de fragilité Gaussien inverse* sur E.

Considérons maintenant le famaux modèle de panne accélérée (AFT) sur E, [voir Bagdonavičius and Nikulin (2002)]. Le modèle AFT est le plus utilisé dans les *essais accélérés*.
Nous disons que la famille $\{S_{x(\cdot)}, x(\cdot) \in E\}$ de fonction de survie sur E constitue le modèle AFT sur E s'il existe une fonction positive $r : E \to \mathbf{R}^1$ et une fonction de survie de base S_0 tels que les l'éléments de la famille $\{S_{x(\cdot)}, x(\cdot) \in E\}$ verifient la relation suivante:

$$S_{x(\cdot)}(t) = S_0\left(\int_0^t r(x(u))du\right), x(\cdot) \in E,$$

où la fonction de survie de base S_0 ne dépond pas de $x(\cdot)$.

Très souvent on suppose que les distributions de base de temps de panne dans le modèle AFT et la loi de la variable de fragilité dans le modèle de fragilité, sont de la famille de lognormale, Gamma, Weibull, et Weibull Généralisée. Dans ces cas la famille des distributions gaussiennes inverses est concurrente à ces 4 familles. Nous étudions les propriétés des modèles de Fragilité et de AFT quand elles sont fondées sur la famille de distributions Gaussiennes inverses.

References

[1] Bagdonavičius, V., Nikukin, M., (2002) accelereted Life Models: Modeling and Statistical analysis. Chapman and Hall.

[2] Chhikara, R.S., Folks, J.L., (1989). The Inverse Gaussian Distribution. Marcel Dekker, New York..

[3] Drost F. (1988), Asymptotics for generalized chi-squared goodness-of-fit tests, amsterdam: Centre for Mathematics and Computer Sciences,CWI Tracs, 48.

[4] Greenwood, P. S. and Nikulin, M. (1996), A guide to chi-squared testing, John Wiley and Sons, New York.

[5] Habib, M.G., Thomas, D.R. (1996). Chi-squared Goodness-of-Fit Tests For Randomly Censored Data. Annals of Statistics, 14, 759-765.

[6] Huber, C., Limnios, N., Mesbah, M., Nikulin, M., (2008) Mathematical Methods in Survival analysis, Reliability and Quality of Life, Willy-ISTE

[7] Seshadri, V. (1993) The Inverse Gaussian Distribution: A Case Study in Exponential Families, Clarendon Press:Oxford,

[8] Seshadri, V. (1999) The inverse Gaussian distribution: statistical theory and applications. Springer,New York

[9] Nikulin M.S.(1973),On a chi-square test for continuous distributions,Thery of probability and applications, 18 , p.638-639.

[10] Nikulin M.S. Chi-square test for continuous distributions with shift and scaie parameters," Teor. Veroyatn.Primen., 18, No. 3, 559-568 (1973).

[11] Voinov, M.,Nikulin,M. (1993), Unbiased Estimators and Their Applications,V.1 Univariate case,Dordrecht: Kluwer Academic Publishers.

[12] Rao, K.C. and Robson, D.S. (1974). A chi-square statistic for goodness-of-fit tests within the exponential family. Commun. Statist. 3 1139-1153