# Prédictions bayésiennes de statistiques d'ordre Avec des données groupées : Cas d'une loi exponentielle

# Bayesian predictions of order statistics with grouped data: The case of an exponential law

Assia Chadli* & Asma Meradji

*Laboratoire LaPS, Département de Mathématiques, Faculté des Sciences*
*Université Badji Mokhtar Annaba, BP 12, 23000, Annaba, Algérie*

*Soumis le : 05.10.2013* *Révisé le : 25.03.2014* *Accepté le : 09.04.2014*

**ملخص**
في نموذج أسي، قمنا بإخضاع n بنود إلى اختبار البقاء على قيد الحياة. نحن مهتمون بالتنبؤ لإحصاءات النظام ووظائفها على عينة مستقبلية المنهج حجمها N من نفس القانون و مستقلة عن العينة المرصودة .المستعمل هو منهج بييز مع قانون التزاوج الطبيعي على معلمة و دالة فقدان من الدرجة الثانية. توجد نتائج في حالة الملاحظات المقطوعة كما نقترح استعمال المعطيات المجمعة ( القسم 2) ثم البيانات من التصميم التجريبي مع التجديد (القسم 3). بالإضافة إلى الكثافة التنبؤية، تم العثور على التنبؤ في كل حالة.

**الكلمات المفتاحية**: تنبؤ بييز- *معطيات مجمعة - الكثافة التنبؤية - المتنبيء*

**Résumé**
Dans un modèle exponentiel, on soumet n items à un test de survie. On s'intéresse à la prédiction de statistiques d'ordre et de fonctions de celles-ci sur un échantillon futur de taille N, issu de la même loi et indépendant de l'échantillon observé. L'approche utilisée est une approche bayésienne avec une loi a priori conjuguée naturelle sur le paramètre et une fonction de perte quadratique. Des résultats existent dans le cas d'observations censurées, aussi, nous nous proposons d'utiliser des données groupées (section 2), puis des données issues d'un plan d'expérience avec renouvellement (section 3). Outre la densité prédictive, un prédicteur a été trouvé dans chaque cas.

**Mots clés :** *Prédiction Bayésienne - Données groupées - Densité prédictive - Prédicteur*

**Abstract**
In an exponential model, n items were subject to a survival test. We are interested in the Bayesian prediction of order statistics and functions of them on a future sample of size N coming from the same law and independent from the observed sample. The used approach is Bayesian, with conjugate prior law on the parameter and a square loss function. The existing results were obtained in the case of censored data. In this work, we propose to use grouped data (section 2), then data from an experimental design with renewal (section 3.) In addition to the predictive density, a predictor was found in each case.

**Keywords:** *Bayesian prediction - Grouped datas - Predictive density - Predictor.*

**Auteur correspondant :* *assiachadli428@hotmail.com*

## 1. INTRODUCTION

Le problème de la prédiction de statistiques d'ordre et de quelques fonctions d'intérêt dépendant de celles-ci a été largement étudié dans des modèles exponentiels. En effet, la loi exponentielle a l'avantage de posséder la propriété de perte de mémoire qui permet d'avoir des calculs aisés et des résultats analytiques simples, donc facilement exploitables. Plusieurs auteurs se sont intéressés à ce problème en utilisant une approche classique, celle du maximum de vraisemblance [1- 4]. L'approche bayesienne a été également utilisée avec des lois à priori conjuguées naturelles ou non informatives des données complètes ou censurées et différentes lois. Bratcher [5] a utilisé des données complètes, Dunsmore [6] a étudié d'une manière générale les tests de survie, Evans et Nigm [7, 8] ont étudié la loi exponentielle tronquée et la loi exponentielle bivariée avec des données censurées de type II; Arnold [9] a considéré la loi de Pareto, Erkanly *et al*. et Van-Batenberg [10, 11] ont appliqué les résultats de la prédiction Bayesienne à la fiabilité. Chadli [12] a considéré le problème de la prédiction de statistiques d'ordre à partir de données censurées de type attribut. Dans l'approche bayesienne, et pour les problèmes de prédictions de statistiques d'ordre et de fonctions de celles-ci, plusieurs types de données ont été utilisées : des données censurées de type II par Evans et Nigm [7, 8] ; des données de type attribut par Chadli [12]. Quant aux données groupées, elles ont été utilisées dans un modèle de Burr par Aludaat *et al*. [14], Amjad et Ayman mais aussi par Makhdoom et Amrollah [15,16], pour des problèmes d'estimations.

Nous nous proposons, dans ce travail, de considérer les problèmes de prédictions de statistiques d'ordre et de fonctions de celles-ci dans un échantillon futur de taille *N*. Les observations utilisées sont issues d'un plan d'expérience avec des données groupées puis, nous considérons le cas d'un plan d'expérience avec renouvellement. Il est supposé que les durées de vie suivent une loi exponentielle à un paramètre dont la densité est donnée par :

$$f(x/\theta)=\theta \exp\{-\theta x\};\theta>0;x>0 \qquad (1)$$

Soit $(X_1, X_2,..., X_n)$ un n-échantillon d'une loi *exp(θ)*. Nous disposons d'une expérience E dans laquelle *n* items sont soumis à un test de survie. Dans les deux plans d'expériences cités plus haut, on s'intéresse à la prédiction dans un échantillon futur $(Y_1,Y_2,...,Y_N)$ indépendant du premier des deux variables d'intérêts suivantes:

$$Y=Y_{(r)};1<r<N \qquad (2)$$

$$Z=\sum_{i=1}^{k}Y_{(i)}+(N-k)Y_{(k)} \qquad (3)$$

Où $(Y_{(1)}, Y_{(2)} \ldots,Y_{(N)})$ représente la statistique d'ordre de $(Y_1,Y_2,...,Y_N)$, *Y* représente la ième statistique d'ordre et Z, la durée de vie globale des N items jusqu'à la kième panne.

Dans le paragraphe 2, outre la densité prédictive, un prédicteur de (2) a été trouvé avec des données groupées. Dans le paragraphe 3 ; la densité prédictive et un prédicteur de (2) et de (3) ont été trouvés dans le cas d'un plan d'expérience avec renouvellement.

## 2. Cas où les données sont groupées

Considérons un plan d'expérience de durée *T* fixée, comprenant à l'origine *n* items soumis à un test de survie. Les durées de vie sont supposées exponentielles de densité donnée en (1). L'intervalle de temps [0, *T*] est subdivisé en *k* intervalles de longueur *T/k*.

Soit $(t_1,t_2,\ldots,t_k)$ tel que $t_j = j\, T/k$ ;

$j= 1,\ldots, k$ des temps d'inspections prédéterminés, avec $t_k=T$ , le dernier instant d'inspection.

On suppose $t_0=0$ et $t_{k+1}\to\infty$.

L'information recueillie à l'issue du test est résumée par les nombres $x_i$ de défaillances ayant eu lieu dans chacun des *k* intervalles $[t_{j-1}, t_j]$ ; $j= 1,\ldots, k$.

$t_0$ $t_1$ …

$x_1$

$t_{k-1}$ $t_k=T$

$x_k$ $n-\sum_{i=1}^{k}x_i$

Soient :

$p_1$= probabilité de panne dans $\left[0,\frac{T}{k}\right]=F_\theta(t_1)$

$p_j$ = probabilité de panne dans $\left]t_{j-1},t_j\right]=F_\theta(t_j)-F_\theta(t_{j-1})$ pour j=1,…, k

En vertu de la propriété de perte de mémoire de la loi exponentielle, les probabilités $p_j$ vérifient la relation de récurrence suivante :

$$p_j = p_1\exp\left\{-\theta(j-1)T/k\right\}$$

et $\sum_{j=1}^{k} p_j = F_\theta(T)$

(4)

En effet

$$p_j = F_\theta(t_j)-F_\theta(t_{j-1})$$
$$=\exp(-\theta t_{j-1})-\exp(-\theta t_j)$$
$$=\exp-(\theta(j-1)T/k)-\exp(-\theta(j)T/k)$$
$$=\exp-(\theta(j-1)T/k)[1-\exp-(\theta T/k)]$$
$$=p_1\exp(-\theta(j-1)T/k)$$

## 2.1 Vraisemblance et densité à posteriori

Soit

$\underline{x}=(x_1,\ldots x_k)$ le vecteur des observations.

On pose,

$$S=\sum_{i=1}^{k}x_i\,;\; S'=nk-\sum_{i=2}^{k}(k-i+1)x_i\,;\; C=\frac{n!}{\prod_{i=1}^{k}x_i!\,(n-s)!}$$

La vraisemblance suit alors une loi multinomiale

$M_{k+1}(n,p_1,\ldots\ldots,p_k)$

$$L(\underline{x}/\theta)=C\prod_{j=1}^{k}p_j^{xj}\left\{1-\sum_{i=1}^{k}p_i\right\}^{n-s}$$

La relation de récurrence (4) nous permet d'écrire la vraisemblance uniquement en fonction de $p_1$:

$$L(\underline{x}/\theta)=Cp_1^S\,exp\left\{-\theta\frac{T}{k}(x_2+2x_3+\ldots\ldots+(k-1)x_k\right\}\left(1-\sum_{i=1}^{k}p_i\right)^{n-S}$$

$$L(\underline{x}/\theta)=C\sum_{i=0}^{S}(-1)^i\binom{S}{i}\exp\left\{-\theta\frac{T}{k}(S'+i)\right\}$$

On suppose sur $\theta$ une distribution à priori de type Gamma, $G(g,h)$, en effet celle-ci est une conjuguée naturelle, la densité a posteriori est alors:

$$p(\theta/\underline{x})=\frac{1}{K\Gamma(g)}\sum_{i=0}^{S}(-1)^i\binom{S}{i}\theta^{g-1}\,exp\left\{-\theta\left[\frac{T}{k}(S'+i)+h\right]\right\}$$

Où

$$K=\sum_{i=0}^{S}(-1)^i\binom{S}{i}\left\{\frac{T}{k}(S'+i)+h\right\}^{-g}$$

## 2.2 Prédiction de *Y = Y(r)*, 1 ≤ r ≤ N

Considérons un échantillon futur de taille N, indépendant de l'échantillon sur lequel a porté l'observation. On s'intéresse alors dans cette section à la prédiction de la r[iéme] statistique d'ordre en utilisant une distribution à priori sur θ de type $G(g,h)$, et une fonction de perte quadratique.

La densité de *Y* est (voir (12)):

$$f(y/\theta)=r\binom{N}{r}\sum_{j=0}^{r-1}(-1)^j\binom{r-1}{j}\theta\,exp\{-\theta y(N-r+1-j)\}$$

Densité prédictive par rapport à une fonction de perte quadratique est :

$$p(y/\underline{x})=\int_0^\infty f(y/\theta)p(\theta/x)d\theta$$

$$=\frac{r\binom{N}{r}}{K\Gamma(g)}\sum_{i=0}^{S}\sum_{j=0}^{r-1}(-1)^{i+j}\binom{S}{i}\binom{r-1}{j}\int_0^\infty\theta^g\,exp\left\{-\theta\left[(N-r+1-j)+H+i\frac{T}{k}\right]\right\}d\theta$$

$$=\frac{gr\binom{N}{r}}{K}\sum_{i=0}^{S}\sum_{j=0}^{r-1}(-1)^{i+j}\binom{S}{i}\binom{r-1}{j}\left\{(N-r+1+j)g+H+i\frac{T}{k}\right\}^{-(g+1)}$$

Où $H=\frac{T}{k}S'+h$

La fonction de répartition prédictive est alors :

$$F(z/\underline{x})=\int_0^z f(y/x)dy$$

$$=\frac{r\binom{N}{r}}{K}\sum_{i=0}^{S}\sum_{j=0}^{r-1}(-1)^{i+j}\binom{S}{i}\binom{r-1}{j}(N-r+1+j)^{-1}$$
$$\times\left\{(H+iT/k)^{-g}-[z(N-r+1+j)+(H+iT/k)]^{-g}\right\}$$

Sous l'hypothèse g >1 ; un prédicteur de $Y_{(r)}$ serait :

$$E(Y/\underline{x})=\frac{r\binom{N}{r}}{(g-1)K}\sum_{i=0}^{S}(-1)^i\binom{S}{i}\{H+iT/k\}^{-(g+1)}\sum_{j=0}^{r-1}(-1)^j\binom{r-1}{j}(N+r+1+j)^{-2}$$

**Cas particulier : r = 1**

Il s'agit de la prédiction de la 1[ére] statistique d'ordre, ce qui se traduit dans un cadre fiabiliste par l'apparition de la 1[ére] panne. La fonction de densité prédictive et la fonction de répartition prédictive sont respectivement données par

$$f(y\ /\underline{x})=\frac{gN}{K}\sum_{i=0}^{S}(-1)^i\binom{S}{i}\left\{Ny+H+i\frac{T}{k}\right\}^{-(g+1)}$$

$$F(z/\underline{x})=1-\frac{1}{K}\sum_{i=0}^{S}(-1)^i\binom{S}{i}[z(N-r+1+j)+H+iT/k]^{-g}$$

## 3. Prédiction de $Y = Y_{(k)}$ quand on dispose d'un test de survie avec renouvellement

On observe durant un temps *T*, *n* items dont les durées de vie sont issues de (1). On suppose de plus que tout item tombé en panne dans [0, *T*] est renouvelé et que le renouvellement est instantané.

L'information disponible à la fin du test est le nombre X de renouvellements ayant eu lieu dans [0, *T*] : On s'intéresse à la prédiction de (2) et (3).

### 3.1 Prédiction de $Y = Y_{(k)}$ ; $1 \le k \le N$

Les intervalles entre deux pannes consécutives suivent une loi expo$(\theta)$ : Le modèle considéré est donc une superposition de processus de Poisson de paramètre $\theta$. On en déduit que X suit une loi de Poisson. Soit $L(x/\theta)$ la vraisemblance :

$$L(x/\theta)=\frac{(\theta nT)^x}{x!}\exp\{-\theta nT\};\ \ \theta>0;\ x=0,1,.$$

Il est clair que la famille de distribution $G(g,h)$ est une famille de conjuguées naturelles pour $\theta$. Soit $p(\theta)$ la densité à priori sur $\theta$ :

$$p(\theta)=\frac{h^g}{\Gamma(g)}\theta^{g-1}\,exp\{-\theta h\};\ \ h,g>0$$

La densité à posteriori sur $\theta$ est alors une loi $G\,(G\,,H\,)$

$$p(\theta/x)=\frac{H^G}{\Gamma(G)}\theta^{G-1}\,exp\,\{-\theta H\}\ ;$$

où $H = k + nT$ et $G = g + x$

Densité prédictive de $Y_{(k)}$

$$p(y/x)=\int_0^\infty p(y/\theta)p(\theta/x)d\theta$$

Où

$$p(y/\theta)=k\binom{N}{k}\sum_{i=0}^{k-1}(-1)^i\binom{k-1}{i}\theta\,exp\{-\theta y(N-k+1+i)\}$$

$$p(y/x)=GH^G k\binom{N}{k}\sum_{i=0}^{k-1}(-1)^i\binom{k-1}{i}[H+y(N-k+1+i)]^{-(G+1)}$$

Fonction de répartition prédictive $F(z/x)$

$$F(z/x)=\int_0^z p(y/x)dy$$

$$GH^G k\binom{N}{k}\sum_{i=0}^{k-1}(-1)^i\binom{k-1}{i}\int_0^z[H+y(N-k+1+i)]^{-(G+1)}dy$$

$$=1-H^G k\binom{N}{k}\sum_{i=0}^{k-1}(-1)^i\binom{k-1}{i}[H+z(N-k+1+i)]^{-G}$$

En particulier pour k = 1

$$F(z_1/x)=1-H^G N(H+N_{z_1})^{-G}$$

**Remarque** : Une connaissance à priori non informative sur $\theta$ correspond à $g,h\to 0$ soit $G=x$ et $H=nT$.

En particulier pour $Y_{(1)}$ ; la fonction de répartition serait :

$$F(z_1/x)=1-N(nT)^x(nT+N_{z_1})^{-x}$$

Un intervalle de prédiction pour $Y_{(1)}$ de la forme (0, z) et de niveau $1-\gamma$ est tel que *z* soit solution de :

$$N\left\{\frac{nT}{nT+Nz}\right\}^x=\gamma \qquad (5)$$

Ce qui nous donne ;

$$z=\frac{nT}{N}\left\{\left(\frac{N}{\gamma}\right)^{1/x}-1\right\} \qquad (6)$$

Prédicteur de $Y_{(K)}$ sous l'hypothèse $G > 1$.

$$E(Y/x)=\frac{H}{G-1}k\binom{N}{k}\sum_{i=0}^{k-1}(-1)^i\binom{k-1}{i}(N-k+1+i)^{-2}$$

Pour $k=1$

$$E(Y_1/x)=\frac{H}{(G-1)N^2} \qquad (7)$$

**3.2 Prédiction de** $Z=\sum_{i=1}^{k}Y_{(i)}+(N-k)Y_{(k)}$

$Z$ suit une loi $G(k,\theta)$ de densité :

$$f(z/\theta)=\frac{\theta^k}{\Gamma(k)}z^{k-1}\exp(-\theta z) \quad z>0; \quad \theta>0$$

***Densité prédictive de Z***

La densité prédictive $p(z/x)$ de $Z$ est :

$$p(z/x)=\frac{H^G}{\Gamma(k)\Gamma(G)}z^{k-1}\int_0^\infty \theta^{k+G-1}\exp\{-\theta(H+z)\}d\theta$$

$$=\frac{H^G}{B(k,G)}\times\frac{z^{k-1}}{(H+z)^{k+G}}; \quad z\geq 0$$

On voit que la loi de $(Z/H)$ sachant que $x$ suit une loi Béta, de type II, de paramètres $G$ et $H$, qu'on peut ramener à une loi de Fisher en faisant la transformation suivante :

$$\frac{G}{kH}Z\to F(2k,2G)$$

***Fonction de répartition prédictive*** $F(z/x)$

$$F(z/x)=\int_0^z p(y/x)dy=I_{z_{H+z}}(k,G)$$

Un intervalle de prédiction de $Z$ de la forme $(0,z)$ et de niveau $(1-\gamma)$ est tel que $z$ soit la solution de

$$I_{z/_{z+H}}(k,G)=1-\gamma \qquad (8)$$

Prédicteur de $Z$ par rapport à une fonction de perte quadratique et sous l'hypothèse $G$>1

$$E(Z/x)=\frac{H^G}{B(k,G)}\int_0^\infty\frac{z^k}{(z+H)^{k+G}}dz$$

$$=H\frac{k}{(G-1)}$$

**Remarque :** pour une connaissance à priori non informative sur $\theta$ correspondant à $g,h\to 0$, un prédicteur de $Z$ serait, sous l'hypothèse $x$ >1.

$$E(Z/x)=k\frac{nT}{x-1} \qquad (9)$$

## 4. CONCLUSION

On obtient des prédicateurs de (2) et (3) de forme analytique très simple et donc facilement exploitables, et ce malgré l'utilisation de plans d'expérience nettement moins contraignants et onéreux que les plans d'expériences complets ou les plans d'expériences avec censures à droite.

## REFERENCES

[1] Dunsmore I.R., 1974. Asymptotic prediction analysis, *Biometrica*, Vol 63, 627-630.

[2] Bancroft G.A. & Dunsmore I.R.,1976. Predictive distribution in life testes under competing causes of failure, *Biometrica*, Vol 63 (1), 195-217.

[3] Prezzi M, Geyskens P & Monteiro P.M., 1996.Reliability approach to survie life prediction of concrete exposed to marine environments, *ACI Materials Journal*, Vol 93 (6), 544-552.

[4] Pan J.N., 1997. Reliability prediction of imperfect servicing systems subject to multiple stress, *Microelectronics and reliability*, Vol 37 (3), 439-445.

[5] Bratcher-Schucany & Hunt, 1971. Bayesian prediction and population size assumptions. *Technometrics*, Vol 13, 678-684.

[6] Dunsmore I.R., 1974. The Bayesian predictive distribution in life testing models, *Technometrics*, Vol 16 (3), 445-450.

[7] Evans I.G. & Nigm A.H.M., 1980. Bayesian prediction for the left truncated exponential distribution, *Technometrics*, Vol 22, (2), 329-340.

[8] Evans I.G. & Nigm A.H.M., 1980. Bayesian 1-sample prediction for the 2-parameter distribution, *I.E.E.E. transaction on reliability*, Vol R-29 (5), 218-229.

[9] Arnold B.C, Bayesian estimation and prediction for Pareto data, 1989. *Journal of the American Statstical Association*, Vol 84, (408) 1079-1084.

[10] Erkanli A., Mazzukhi T.A. & Soyer R, 1998. Bayesian computation for a class of reliability growth models, *Technometrics*, Vol 40 (1), 14-23.

[11] Prezzi M, Geyskens P & Monteiro P.M., 1996. Reliability approach to survives life prediction of concrete exposed to marine environments, *ACI Materials Journal*, Vol 93 (6), 544-552.

[12] Chadli A, 2004. Prédictions Bayésiennes en fiabilité: Cas d'une loi exponentielle, *Publications de l'institut de statistiques de Paris. XXXXVIII*, fasc. 1-2, 3-17.

[13] Kundu D. & Raqab M.Z., 2012. Bayesian inference and prediction of order statistics for a typeII censored Weibull distribution, *Journal of statistical planning and inference*, 142, 41-47.

[14] Aludaat, K.M., Alodat M.T. & Alodat T.T., 2008. Parameter estimation of Burr type X distribution for grouped data. *Journal of Applied Mathematical Sciences,* 2 (9), 415-423.

[15] Amjad A. & Ayman B., 2006. Interval estimation for the scale parameter of Burr type X distribution based on grouped data. *Journal of Modern Applied Statistical Methods* 3, 386-398.

[16] Makhdoom I. & Jafari A., 2011. Bayesian Estimations on the Burr Type XII Distribution using grouped and un-grouped data. *Australian Journal of Basic & Applied Sciences Vol.* 5(6), 1525-1531.

[17] Alodat M., Alaudaat K. & Alodat K., 2008. Bayesian Prediction Interval from Grouped Data: Exponential Distribution *abhathal-yarmouk,* "*Basic Science and Engineering*" Vol. 17 (2), 517-524.