C. R. Acad. Sci. Paris, Ser. I 335 (2002) 557–560

Statistique/*Statistics*
(Probabilités/*Probability Theory*)

Mesures prédictives dans le modèle de Cox–Dirichlet à censures aléatoires

Fatiha Messaci

Département de mathématiques, Université de Constantine, BP 325, route Ain El Bey, 25017 Constantine, Algérie

Reçu le 12 mars 2002 ; accepté le 28 juin 2002

Note présentée par Paul Deheuvels.

Résumé Dans ce travail nous étendons au cas d'observations aléatoirement censurées l'étude des mesures prédictives dans le modèle non paramétrique bayésien de Cox–Dirichlet effectuée, pour des observations non censurées, par N. Gouget et J.P. Raoult, dans [3]. Nous montrons que dans les différents cas de censure envisagés, les mesures prédictives restent régulières par portions au sens de l'article [3]. Il en résulte une partition de l'espace des observations en parties contenues dans des sous espaces vectoriels, sur chacune desquelles la mesure prédictive est absolument continue par rapport à la mesure de Lebesgue. Nous calculons, dans le cas de la censure à droite, les densités ainsi définies. ***Pour citer cet article : F. Messaci, C. R. Acad. Sci. Paris, Ser. I 335 (2002) 557–560.***
© 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

Predictive measures in the Cox–Dirichlet model with random censures

Abstract This paper is devoted to the extension to randomly censored observations of the study of predictive measures in the nonparametric bayesian Cox–Dirichlet model, which has been developed for non censored observations by N. Gouget and J.P. Raoult in [3]. We show that in all cases, the predictive measures stay piecewise regular (as defined in [3]). This implies that there exists a partition of the space of observations in subsets, included in linear subspaces, such that in each of them the predictive measure is absolutely continuous w.r.t. the Lebesgue measure. We compute the so defined densities for right censoring. ***To cite this article: F. Messaci, C. R. Acad. Sci. Paris, Ser. I 335 (2002) 557–560.***
© 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

Abridged English version

A probability measure on $\mathbb{R}^n$ is said piecewise regular (see [3]) if, for any permutation $\sigma = (\sigma_1, \ldots, \sigma_n)$ the restriction of its survival function (s.f.) S to

$$\bar{\Delta}_\sigma = \left\{(t_i)_{1\leqslant i\leqslant n} : \forall j \in \{1, \ldots, n-1\},\ t_{\sigma_j} \leqslant t_{\sigma_{j+1}}\right\}$$

admits a continuous n-derivative $\partial^n S_\sigma / \partial t_1 \cdots \partial t_n$.

It is known that a piecewise regular measure is not necessarily absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}^n$ but that absolute continuity holds for its restrictions to any l dimensional (with $1 \leqslant l \leqslant n$) subset defined by equalities and strict inequalities between coordinates, that is any

$$\Delta_\sigma^\lambda = \left\{(t_i)_{1\leqslant i\leqslant n} : \forall j \in \{1, \ldots, l\},\ t_{\sigma_{n_{j-1}+1}} = \cdots = t_{\sigma_{n_j}},\ \forall j \in \{1, \ldots, l-1\},\ t_{\sigma_{n_j}} < t_{\sigma_{n_{j+1}}}\right\},$$

Adresse e-mail : f_messaci@yahoo.fr (F. Messaci).

© 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS. Tous droits réservés
S1631-073X(02)02510-4/FLA

where $\lambda = (\lambda_1, \ldots, \lambda_l)$ with $\sum_{j=1}^{l} \lambda_j = n$ and $n_j = \sum_{h=1}^{j} \lambda_h$ (with $n_0 = 0$).

This property holds for the predictive measure in the following bayesian Cox–Dirichlet model which provides a modelisation for life duration data with multiple right censoring: we observe n independent random variables, i.e.

$$T_i = \inf(T_{i,1}, \ldots, T_{i,m}) = T_{i,1} \wedge \ldots \wedge T_{i,m}$$

$(1 \leqslant i \leqslant n)$ where $T_{i,1}, \ldots, T_{i,m}$ are independent and, for each k $(1 \leqslant k \leqslant m)$, there exists a baseline s.f. G_k and strictly positive fixed constants $c_{i,k}$ so that the s.f. of $T_{i,k}$ is $G_k^{c_{i,k}}$. For each k, G_k is unknown and endowed with a Dirichlet prior (as defined in [1]), characterized by a positive finite measure α_k^*; the priors relative to different values of k are independent. The case with $m = 1$ had been considered in [3]; by generalizing the computations in that paper we obtain a closed form for the density of the predictive measure on each subset Δ_σ^λ w.r.t. the Lebesgue measure (see Theorem 1). Piecewise regularity also holds for left censoring (observation of $T_{i,1} \vee T_{i,2}$) and for every form of right–left censoring (observation of $[T_{i,1} \wedge (T_{i,2} \vee T_{i,3})] \vee (T_{i,2} \wedge T_{i,3})$, $(T_{i,1} \vee T_{i,3}) \wedge (T_{i,2} \vee T_{i,3})$ or $[T_{i,1} \vee (T_{i,2} \wedge T_{i,3})] \wedge T_{i,2}$) where computation of densities is feasable but intricated. Details of proofs and computations are given in [4].

1. Introduction

Dans [3], Gouget et Raoult introduisent la notion de mesure régulière par portions sur $\mathbb{R}^n$ et, partant de la fonction de survie de la loi prédictive dans le modèle bayésien de Cox–Dirichlet, calculée dans [5], ils montrent que la mesure qui lui est associée rentre dans ce cadre. Ils en déduisent les densités restreintes de la loi prédictive à des sous ensembles caractérisés par des égalités et des inégalités strictes entre les observations. Pour préciser ce résultat, rappelons les notions et notations suivantes de [3].

Une configuration d'ex æquo est toute suite de nombres entiers strictement positifs $\lambda = (\lambda_1, \ldots, \lambda_l)$ telle que $\sum_{j=1}^{l} \lambda_j = n$. Nous notons $n_j = \sum_{h=1}^{j} \lambda_h$ (avec $n_0 = 0$). Pour chaque n-permutation $\sigma = (\sigma_1, \ldots, \sigma_n)$ et chaque configuration d'ex æquo λ nous introduisons le sous ensemble de $\mathbb{R}^n$ défini par

$$\Delta_\sigma^\lambda = \left\{ (t_i)_{1 \leqslant i \leqslant n} : \forall j \in \{1, \ldots, l\},\ t_{\sigma_{n_{j-1}+1}} = \cdots = t_{\sigma_{n_j}},\ \forall j \in \{1, \ldots, l-1\},\ t_{\sigma_{n_j}} < t_{\sigma_{n_j+1}} \right\}.$$

Δ_σ^λ est l'ensemble de toutes les suites dans $\mathbb{R}^n$ qui peuvent être réordonnées en croissant par l'usage de la permutation σ et dans lesquelles il y a exactement l blocs d'ex æquo qui, lorsqu'on les recense en croissant, ont les tailles successives $\lambda_1, \ldots, \lambda_l$.

Δ_σ^λ est en bijection avec un sous ensemble de $\mathbb{R}^l$ par l'application : $(t_i)_{1 \leqslant i \leqslant n} \mapsto (t_{\sigma_{n_j}})_{1 \leqslant j \leqslant l}$, ce qui permet de définir la mesure de Lebesgue sur Δ_σ^λ par transport de la mesure de Lebesgue sur $\mathbb{R}^l$.

Pour $\lambda = (1, \ldots, 1)$, la suite de longueur n dont tous les éléments sont égaux à 1, nous obtenons les sous ensembles $\Delta_\sigma = \{(t_i)_{1 \leqslant i \leqslant n} : \forall j \in \{1, \ldots, n-1\},\ t_{\sigma_j} < t_{\sigma_{j+1}}\}$, dont la fermeture est $\bar{\Delta}_\sigma = \{(t_i)_{1 \leqslant i \leqslant n} : \forall j \in \{1, \ldots, n-1\},\ t_{\sigma_j} \leqslant t_{\sigma_{j+1}}\}$.

Une mesure finie sur $\mathbb{R}^n$ est dite régulière par portions si pour toute n-permutation σ, la restriction de sa fonction de survie, S, à $\bar{\Delta}_\sigma$ notée S_σ, admet la n-dérivée continue $\partial^n S_\sigma / \partial t_1 \cdots \partial t_n$.

Dans ce travail, dont une version plus détaillée est [4], nous étendons le résultat de [3] au cas où les observations sont censurées soit à droite, soit à gauche, soit en même temps à droite et à gauche (censure mixte). Pour tous ces cas nous montrons que la mesure prédictive obtenue reste régulière par portions. Les densités prédictives restreintes aux Δ_σ^λ sont calculées, dans le cas de la censure à droite.

Dans toute la suite, pour tout i $(1 \leqslant i \leqslant n)$, l'observation T_i résulte (en un sens que nous préciserons dans chaque cas) de m variables aléatoires inobservables $T_{i,k}$ $(1 \leqslant k \leqslant m)$ strictement positives indépendantes et, pour tout k, le vecteur aléatoire $(T_{i,k})_{1 \leqslant i \leqslant n}$ est lui même régi par le modèle bayésien de Cox–Dirichlet étudié dans [3]. De manière précise, il existe une famille de fonctions de survie de base $(G_k)_{1 \leqslant k \leqslant m}$ et une famille doublement indexée de contraintes strictement positives $c = (c_{i,k})_{1 \leqslant i \leqslant n,\ 1 \leqslant k \leqslant m}$ de telle sorte que la fonction de survie de $T_{i,k}$ soit la puissance $G_k^{c_{i,k}}$. D'autre part chaque famille de lois de base est munie d'un a priori de Dirichlet $\mathcal{D}_{\alpha_k^*}$, caractérisé par une mesure finie α_k^* sur $\mathbb{R}_+$, de fonction de survie α_k (i.e. $\alpha_k(t) = \alpha_k^*(]t, +\infty[)$. Rappelons (voir par exemple [2]) que cela signifie que pour tout k et pour toute suite strictement croissante $(u_0, u_1, \ldots, u_s, u_{s+1})$ où $u_0 = 0$ et $u_{s+1} = +\infty$, la suite aléatoire $(1 - G_k(u_1), G_k(u_1) - G_k(u_2), \ldots, G_k(u_{s-1}) - G_k(u_s))$ admet pour densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^s$ l'application :

$$(y_1, \ldots, y_s) \longrightarrow \frac{\Gamma(\alpha_{0,k})}{\Gamma(\alpha_{1,k}) \ldots \Gamma(\alpha_{s+1,k})} \prod_{i=1}^{s} y_i^{\alpha_{i,k}-1} \times \left(1 - \sum_{i=1}^{s} y_i\right)^{\alpha_{s+1,k}-1} 1_{As}(y_1, \ldots, y_s),$$

où Γ désigne la fonction gamma eulérienne, $A_s = \{(y_i)_{1 \leqslant i \leqslant s} : \forall i \;\; 0 < y_i < 1, \;\; \sum_{i=1}^{s} y_i < 1\}$ et $\forall i \in \{1, \ldots, s+1\}$, $\alpha_{i,k} = \alpha_k^*(]u_{i-1}, u_i])$, $\alpha_{0,k} = \sum_{i=1}^{s+1} \alpha_{i,k}$.

Nous supposons enfin que les m fonctions de survie de base sont a priori indépendantes, autrement dit $(G_k)_{1 \leqslant k \leqslant m}$ est muni de l'a priori $\prod_{k=1}^{m} \mathcal{D}_{\alpha_k^*}$.

2. Censure à droite

Ici

$$T_i = \inf(T_{i,1}, \ldots, T_{i,m}) = T_{i,1} \wedge \cdots \wedge T_{i,m}.$$

Ce modèle peut en particulier être utilisé si les variables $T_{i,1}$ sont des durées de vie indépendantes, subissant chacune $m - 1$ censures à droite $T_{i,2}, \ldots, T_{i,m}$. C'est pourquoi nous l'appelerons modèle bayésien de Cox–Dirichlet à censures à droite multiples indépendantes. L'expression de la fonction de survie permet de voir que la mesure prédictive est régulière par portions. Ensuite, par application du Théorème 1 de [3], nous déduisons les densités correspondantes données par le théorème suivant.

THÉORÈME 1. – *Soit μ la loi prédictive de n durées de survie dans le modèle bayésien de Cox–Dirichlet à censures à droite multiples et indépendantes. Les fonctions de survie des mesures de Dirichlet sont supposées continûment dérivables. Soit $\lambda = (\lambda_j)_{1 \leqslant j \leqslant l}$ une configuration d'ex æquo, selon laquelle nous décomposons chacune des m suites de contraintes $(c_{i,k})_{1 \leqslant i \leqslant n}$ en l sous suites successives $(c_k^j)_{1 \leqslant j \leqslant l}$ avec $c_k^j = (c_{\upsilon,k}^j)_{1 \leqslant \upsilon \leqslant \lambda_j}$; nous introduisons les sommes partielles supérieures de contraintes $e_{j,k} = \sum_{h=j}^{l} \sum_{\upsilon=1}^{\lambda_h} c_{\upsilon,k}^h$.*

Alors la restriction de μ à Δ_σ^λ admet comme densité par rapport à la mesure de Lebesgue la fonction g_σ^λ définie par :

$$g_\sigma^\lambda(t_1, \ldots, t_n) = (-1)^l \prod_{k=1}^{m} \frac{\Gamma(\alpha_{0,k})}{\Gamma(\alpha_{0,k} + e_{1,k})} \prod_{j=1}^{l} B_j(x_j) C_j(x_j),$$

où x_j est la valeur commune des observations t_i appartenant au j^{ieme} bloc d'ex æquo, une fois rangées en ordre croissant, $B_j(x_j) = \prod_{k=1}^{m} \frac{\Gamma(\alpha_k(x_j)+e_{j,k})}{\Gamma(\alpha_k(x_j)+e_{j+1,k})}$ et, $P_{\lambda_j - 1}$ désignant l'ensemble de tous les sous ensembles de $\{1, \ldots, \lambda_j - 1\}$,

$$C_j(x_j) = \sum_{M \in P_{\lambda_j}-1} (-1)^{\operatorname{Card} M} \sum_{k=1}^{m} \alpha_k'(x_j)$$
$$\times \left[(\operatorname{Ln}\Gamma)' \left(\alpha_k(x_j) + e_{j+1,k} + c_{\lambda_j,k}^j + \sum_{\upsilon \in M} c_{\upsilon,k}^j \right) - (\operatorname{Ln}\Gamma)' \left(\alpha_k(x_j) + e_{j+1,k} + \sum_{\upsilon \in M} c_{\upsilon,k}^j \right) \right].$$

3. Censure à gauche

Soit ici $m = 2$ et

$$T_i = \sup(T_{i,1}, T_{i,2}) = T_{i,1} \vee T_{i,2}.$$

Le calcul de la fonction de survie prédictive (voir [4], 4) permet de voir que la mesure qui lui est associée est régulière par portions. L'expression de la densité est plus lourde que pour la censure à droite ; par exemple, sur la diagonale dans le cas $n = 2$ c'est :

$$\begin{aligned}
x \longrightarrow\ & a\frac{A_c(x)}{A_0(x)}\big[\ln A_{c_2}(x) - \ln A_c(x) - \ln A_0(x) + \ln A_{c_1}(x)\big] \\
&+ b\frac{B_d(x)}{B_0(x)}\big[\ln B_{d_2}(x) - \ln B_d(x) - \ln B_0(x) + \ln B_{d_1}(x)\big] \\
&+ a\frac{A_c(x)}{A_0(x)} b\frac{B_d(x)}{B_0(x)}\big[\ln A_{c_2}(x) - \ln A_0(x) + \ln B_{d_2}(x) - \ln B_0(x) \\
&\qquad\qquad - \ln A_c(x) + \ln A_{c_1}(x) - \ln B_d(x) + \ln B_{d_1}(x)\big]' \\
&+ a\frac{A_c(x)}{A_0(x)}\frac{B_0(0)}{B_0(x)}\left(\frac{B_{d_2}(x)}{B_{d_2}(0)} + \frac{B_{d_1}(x)}{B_{d_1}(0)}\right)\big[\ln A_c(x) - \ln A_{c_1}(x) - \ln A_{c_2}(x) + \ln A_0(x)\big]' \\
&+ b\frac{B_d(x)}{B_0(x)}\frac{A_0(0)}{A_0(x)}\left(\frac{A_{c_2}(x)}{A_{c_2}(0)} + \frac{A_{c_1}(x)}{A_{c_1}(0)}\right)\big[\ln B_d(x) - \ln B_{d_1}(x) - \ln B_{d_2}(x) + \ln B_0(x)\big]',
\end{aligned}$$

où $c = c_{1,1} + c_{2,1}$, $d = c_{1,2} + c_{2,2}$, $a = \frac{\Gamma(\alpha_1(0))}{\Gamma(\alpha_1(0)+c)}$, $b = \frac{\Gamma(\alpha_2(0))}{\Gamma(\alpha_2(0)+d\,)}$, $A_h(t) = \Gamma(\alpha_1(t) + h)$ et $B_h(t) = \Gamma(\alpha_2(t) + h)$.

4. Censures mixtes

Nous nous intéressons aux trois modèles suivants.

Modèle symétrique. Ici l'observation d'indice i est :

$$T_i = \big[T_{i,1} \wedge (T_{i,2} \vee T_{i,3})\big] \vee (T_{i,2} \wedge T_{i,3}).$$

Modèle à censure initiale à droite par T_2. Ici l'observation d'indice i est :

$$T_i = (T_{i,1} \vee T_{i,3}) \wedge (T_{i,2} \vee T_{i,3}).$$

Modèle à censure initiale à gauche par T_3. Ici l'observation d'indice i est :

$$T_i = \big[T_{i,1} \vee (T_{i,2} \wedge T_{i,3})\big] \wedge T_{i,2}.$$

Pour tous ces cas, le calcul de la fonction de survie (voir [4], 5) permet de voir que la mesure prédictive est régulière par portions.

Références bibliographiques

[1] T.S. Ferguson, A Bayesian analysis of some nonparametric problems, Ann. Statist. 1 (1973) 209–230.
[2] N. Gouget, Statistique semi-paramétrique Bayésienne de durées de vie. Résultats théoriques et mise en œuvre en fiabilité industrielle, Thèse, Université de Marne-la-Vallée, France, 1999.
[3] N. Gouget, J.P. Raoult, Computation of predictive densities in the semi-parametric Bayesian Cox–Dirichlet model, Nonparametric Statist. 10 (1999) 307–341.
[4] F. Messaci, J.P. Raoult, Etude des mesures prédictives dans le modèle de Cox–Dirichlet à censures aléatoires, Prépublication de l'Équipe d'Analyse et de Mathématiques Appliquées de l'Université de Marne-La-Vallée, 14/2001, France.
[5] M. Ruggiero, Bayesian semiparametric estimation of proportional hazard model, J. Econometrics 62 (1994) 272–300.