Available online at www.sciencedirect.com
ScienceDirect
C. R. Acad. Sci. Paris, Ser. I 347 (2009) 305–308
COMPTES RENDUS MATHEMATIQUE

Statistics

On the quantile process under progressive censoring

Sergio Alvarez-Andrade

Laboratoire de mathématiques appliquées, Université de technologie de Compiègne, B.P. 529, 60205 Compiègne cedex, France

Received 4 July 2008; accepted after revision 22 January 2009

Available online 14 February 2009

Presented by Paul Malliavin

Abstract

This work deals with an asymptotic almost-sure representation of the quantile process under type-II progressive censoring. A convergence rate of the law-of-the-iterated-logarithm type is obtained for this representation. ***To cite this article: S. Alvarez-Andrade, C. R. Acad. Sci. Paris, Ser. I 347 (2009).***

Résumé

Sur le processus quantile sous censure progressive. Nous travaillons sur une représentation asymptotique presque-sûre du processus quantile sous censure progressive du type-II. Nous obtenons pour cette représentation une vitesse de convergence de type loi du logarithme itéré (LIL). ***Pour citer cet article : S. Alvarez-Andrade, C. R. Acad. Sci. Paris, Ser. I 347 (2009).***

Version française abrégée

Contexte : On considère n variables aléatoires $X_1, \dots, X_n$ indépendantes et identiquement distribuées (i.i.d.) pouvant représenter la durée de vie de n matériels en marche. Lors de la première défaillance observée notée $X_{1:m:n}$, on prélève au hasard r_1 matériels parmi les $n-1$ matériels restant en fonctionnement de sorte que seulement $n-1-r_1$ matériels demeurent sous observation. Lors de la deuxième défaillance observée notée $X_{2:m:n}$, on prélève au hasard r_2 matériels parmi les $n-2-r_1$ restant en fonctionnement de sorte que seulement $n-2-r_1-r_2$ matériels demeurent sous observation. Lors de la m-ième défaillance notée $X_{m:m:n}$, les $r_m = n-m-r_1-\cdots-r_{m-1}$ matériels restant en fonctionnement sont censurés.

Notre objectif est d'étudier le comportement asymptotique du processus quantile $(X_{[\alpha m]:m:n})_{\alpha\in[0,a]}$ pour $(0 < a < 1)$. Nous utilisons les résultats du type LIL établis dans (cf. [8] et [6]) pour le tableau triangulaire pondéré apparaissant dans la représentation presque-sûre donnée dans [2]. Nous en déduisons une représentation asymptotique presque-sûre du processus quantile ainsi que la vitesse de convergence associée. Nous raffinons ainsi les résultats établis dans [1].

E-mail address: sergio.alvarez@utc.fr.

1631-073X/$ – see front matter
doi:10.1016/j.crma.2009.01.019

1. Introduction

Let $X_1, \dots, X_n$ be independent and identically distributed (i.i.d.) random variables standing for the lifetimes of n items. A progressive type-II-right-censored sample may be obtained in the following way. At the first failure time, denoted by $X_{1:m:n}$, a prespecified number r_1 of the $n-1$ remaining items are selected at random and removed, so that only $n-1-r_1$ items remain under observation. At the second failure time denoted by $X_{2:m:n}$, a prespecified number r_2 of the $n-2-r_1$ remaining items are selected at random and removed, so that only $n-2-r_1-r_2$ items remain under observation and so on. At the mth failure time denoted by $X_{m:m:n}$, all the remaining $r_m = n-m-r_1-\cdots-r_{m-1}$ remaining items are censored. Therefore, a progressive type-II right-censoring scheme is specified by integers n, m, and $r_1, \dots, r_m$ with the constraints $n-m-r_1-\cdots-r_{m-1} \geqslant 0$ and $1 \leqslant m \leqslant n$. From now on, we assume that the X_i's have common distribution function F with density f. The survival function, also called the reliability function, is denoted by $R = 1-F$. We denote by $\lambda = f/R$ the hazard rate function and $\Lambda = \int_0^{\cdot} \mathrm{d}F/R = -\ln(R)$ the cumulative hazard rate function.

Our purpose is to study the asymptotic behavior of the quantile process $(X_{[\alpha m]:m:n})_{\alpha\in[0,a]}$ $(0<a<1)$ as $m \to \infty$. Specifically, we obtain an almost-sure representation with the associated convergence rate. This refines the results obtained in [1]. Our almost-sure representation with the associated convergence rate holds under the following assumptions:

A1. $\alpha \in [0,a]$ and $0<a<1$;
A2. $(r_i)_{i\geqslant 1}$ is a bounded sequence of non-negative integers ($r_i \leqslant K$ for all $i \geqslant 1$);
A3. $\bar{r} = m^{-1}\sum_{i=1}^{m} r_i = r + o(m^{-1/2})$ where r is a non-negative real number.
A4. Let $G = 1-(1-F)^{r+1}$. There exists a real number $\varepsilon \in [0,a)$ such that $G^{-1}([\varepsilon,a]) \subset (c,b) \subset \mathbb{R}^+$ and λ is continuous and strictly positive on (c,b).

2. Result and proofs

This section is devoted to the statement and the proof of our approximation result. The main ingredient of the proof consists of using the almost-sure representation of $\Lambda(X_{i:m:n})$ derived in [2] and then applying the LIL for triangular arrays. The mentioned almost-sure representation of $\Lambda(X_{i:m:n})$ can be stated as follows. There exists a triangular array $(Z_j^{(m)})_{1\leqslant j\leqslant m}$ of i.i.d. exponentially distributed random variables with mean 1 (denoted by $\varepsilon(1)$), such that the equality below holds almost-surely:

$$\Lambda(X_{i:m:n}) = \sum_{j=1}^{i} \frac{Z_j^{(m)}}{\alpha_j^m}, \tag{1}$$

where $(\alpha_j^m)_{1\leqslant j\leqslant m}$ is a triangular array of non-negative integers defined for $1 \leqslant j \leqslant m$ by

$$\alpha_j^m = r_j + \cdots + r_m + m - j + 1.$$

Let us introduce some additional notations. Let $u_\alpha = G^{-1}(\alpha)$ be the α-quantile of the distribution function G where G^{-1} is taken in the generalized inverse sense when it is not invertible. We note $\log(t) = \log(t \vee e)$ and $\log_2(t)$ the two-iterated logarithm.

Hereafter, we expose our approximation result for the quantile process. Without loss of generality, we will assume that all the random variables and the stochastic processes introduced throughout are defined on the same probability space. By [8] this space is well defined (see also [7] and references therein).

Proposition 2.1. *Under conditions* A1–A4*, we have*:

$$\lambda(u_\alpha)\sqrt{m}(X_{[\alpha m]:m:n} - u_\alpha) = \frac{B_m(\alpha)}{(r+1)(1-\alpha)} + \mathrm{O}_{a.s.}\left(\frac{\log(m)}{\sqrt{m}}\right), \tag{2}$$

where $\{B_m(t),\ 0\leqslant t\leqslant 1\}_{m\geqslant 1}$ *is a sequence of Brownian bridges. Moreover*

$$\lambda(u_\alpha)\sqrt{m}(X_{[\alpha m]:m:n} - u_\alpha) = \frac{W(\alpha m)}{(r+1)\sqrt{m(1-\alpha)}} + \mathrm{O}_{a.s.}(\delta_m), \tag{3}$$

where $\delta_m = (\log_2(m)/m)^{1/4}(\log(m))^{1/2}$ *and* $\{W(t)\}_{t\geqslant 0}$ *is a standard Wiener process.*

Proof. By (1) and a Taylor expansion, we have

$$X_{[\alpha m]:m:n} \overset{a.s.}{=} \Lambda^{-1}\big(\Lambda(u_\alpha)\big) + \frac{\beta_{[\alpha m]}}{\lambda(\Lambda^{-1}(\Lambda(\phi_m)))}, \tag{4}$$

where $\beta_{[\alpha m]} = \sum_{j=1}^{[\alpha m]} Z_j^{(m)}/\alpha_j^m - \Lambda(u_\alpha)$ and ϕ_m is such that $\Lambda(\phi_m)$ belongs to the segment with extremities $\Lambda(u_\alpha)$ and $\Lambda(u_\alpha) + \beta_{[\alpha m]}$. To study $\beta_{[\alpha m]}$, we write for $\epsilon > 0$ that $\sqrt{m}\beta_{[\alpha m]} = I^{(m)}(\epsilon) + II^{(m)} + III^{(m)}$, with

$$I^{(m)}(\epsilon) = \frac{1}{m^{1/2+\varepsilon}}\left(\sum_{j=1}^{[\alpha m]} a_{[\alpha m],\, j}\big(Z_j^{(m)} - 1\big)\right), \quad \text{for } \varepsilon > 0,$$

$$II^{(m)} = \sqrt{m}\sum_{j=1}^{[\alpha m]}\left(\frac{1}{\alpha_j^m} - \frac{1}{(r+1)(m-j+1)}\right), \qquad III^{(m)} = \sqrt{m}\left(\sum_{j=1}^{[\alpha m]} \frac{Z_j^{(m)}}{(r+1)(m-j+1)} - \Lambda(u_\alpha)\right),$$

where the weights in $I^{(m)}$ are given by $a_{[\alpha m],j} = m^{1+\varepsilon}(\frac{1}{\alpha_j^m} - \frac{1}{(r+1)(m-j+1)})$.

To study the asymptotic behavior of the term $I^{(m)}(\epsilon)$, we outline the following points:

(i) The first step consists of replacing the triangular array $(Z_j^{(m)})$ in (1) by a sequence of i.i.d. random variables. Mimicking [7], we let $u_0 = 0$, $u_m = [\alpha m]$ and set $Z_{u_{m-1}+j} = Z_j^{(m)}$, $1 \leqslant j \leqslant [\alpha m]$, $m \geqslant 1$. Then, the variables $\{Z_{u_{m-1}+j},\ 1 \leqslant j \leqslant [\alpha m],\ m \geqslant 1\}$ constitute an i.i.d. sequence of $\varepsilon(1)$ random variables. With these arguments, we can replace $\sum_{j=1}^{[\alpha m]} Z_j^{(m)}/\alpha_j^{(m)}$ by $\sum_{j=1}^{[\alpha m]} Z_{u_{m-1}+j}/\alpha_{u_{m-1}+j}$ where $\alpha_{u_{m-1}+j} = \alpha_j^{(m)}$. This argument is also valid for the term $III^{(m)}$.

(ii) We can use Theorem 1 and Remark 7 of [8] by setting their function $B(u)$ to $B(u) = u^{1/2+\varepsilon}$, for $\varepsilon > 0$ such that $1/2 + \varepsilon < 1$.

(iii) For $0 < \alpha \leqslant a$, we have

$$\sup_{1\leqslant j\leqslant[\alpha m]} |a_{[\alpha m],j}| = \sup_{1\leqslant j\leqslant[\alpha m]} m^{1+\epsilon}\left|\frac{(r+1)(m-j+1) - \alpha_j^m}{\alpha_j^m(r+1)(m-j+1)}\right|$$
$$\leqslant \frac{1}{m^{1-\epsilon}(r+1)(1-\alpha)} \sup_{1\leqslant j\leqslant[\alpha m]} \big|(r+1)(m-j+1) - \alpha_j^m\big|.$$

This jointly with

$$\sup_{1\leqslant j\leqslant[\alpha m]}\left|\frac{\alpha_j^m}{m-j+1} - (r+1)\right| = \mathrm{o}\big(m^{-1/2}\big),$$

under A1–A3, which is an extension of Lemma 4.1 of [3], gives $\sup_{1\leqslant j\leqslant[\alpha m]} |a_{[\alpha m],j}| = \mathrm{O}(m^{-1/2})$.

From (i)–(iii) and Theorem 1 of [8], we get $I^{(m)} \to 0$ a.s. Remark that we can extend this result to $\alpha = 0$. From Corollary 1 in [1], we get $II^{(m)} = \mathrm{o}(1)$.

For the remaining term $III^{(m)}$ (recall (i)), let us remark that, denoting by U an exponentially distributed random variable with mean $r+1$ and distribution function F_U, we have:

$$\Lambda(u_\alpha) = F_U^{-1}(\alpha) = \frac{1}{r+1}\log\left(\frac{1}{1-\alpha}\right) \quad \text{and} \quad f_u\big(F_U^{-1}(\alpha)\big) = (1-\alpha)(r+1).$$

Moreover, if $Q_m(\alpha)$ denotes the $[\alpha m]$th order statistic of m i.i.d. random variables with exponential law of mean $r+1$, we have the following representation: $III^{(m)} \overset{d}{=} \sqrt{m}(Q_m(\alpha) - F_U^{-1}(\alpha))$, where $\overset{d}{=}$ denotes the equality in distribution. By this representation and by [5] p. 143 (see also [4]), on the behavior of $q_m(\cdot) = \sqrt{m}\,(Q_m(\cdot) - F_U^{-1}(\cdot))$, we get on the one hand

$$III^{(m)} = \frac{B_m(\alpha)}{(1-\alpha)(r+1)} + \mathrm{O}_{a.s}\left(\frac{\log m}{\sqrt{m}}\right),$$

which (together with results on $I^{(m)}$ and $II^{(m)}$) gives (2). On the other hand, we have

$$III^{(m)} = \frac{K(\alpha, m)}{(1-\alpha)(r+1)\sqrt{m}} + \mathrm{O}_{a.s}\left(\left(\frac{\log_2(m)}{m}\right)^{1/4}\big(\log(m)\big)^{1/2}\right),$$

where $K(\alpha, m)$ is a Kiefer process (see [5]). Since $W(y) = (x_0(1-x_0))^{-1/2}K(x_0, y)$ for $0 < x_0 < 1$ is a Wiener process, we get:

$$III^{(m)} = \frac{W(\alpha m)}{(r+1)\sqrt{(1-\alpha)m}} + \mathrm{O}_{a.s}\left(\left(\frac{\log_2(m)}{m}\right)^{1/4}\big(\log(m)\big)^{1/2}\right),$$

and it ends (once again with results on $I^{(m)}$ and $II^{(m)}$) the proof of (3).

Acknowledgements

I am indebted to D. Li for giving me the Li and Tomkins (2001) preprint, and to J. Cuzick who informed me about the Li and Tomkins result. The author is grateful to anonymous referees for carefully reading the paper and for their comments which greatly improved the presentation of this paper.

References

[1] S. Alvarez-Andrade, L. Bordes, Empirical quantile process under type-2 progressive censoring, Statist. Probab. Lett. 68 (2004) 111–123.
[2] N. Balakrishnan, R. Aggarwala, Progressive Censoring, Birkhäuser, 2004.
[3] L. Bordes, Non-parametric estimation under progressive censoring, J. Statist. Plann. Inference 119 (2004) 171–189.
[4] M. Csörgő, L. Horváth, Weighted Approximations in Probability and Statistics, John Wiley and Sons, England, 1993.
[5] M. Csörgő, P. Révész, Strong Approximations in Probability and Statistics, Academic Press, New York, 1981.
[6] J. Cuzick, A strong law for weighted sums of i.i.d. random variables, J. Theoret. Probab. 8 (1995) 625–641, Erratum, "A strong law for weighted sums of i.i.d. random variables", J. Theoret. Probab. 14 (2001) 605.
[7] D. Li, M.L. Huang, A. Rosalsky, Strong invariance principles for arrays, Bull. Inst. Math. Acad. Sinica 28 (2000) 167–181.
[8] D. Li, R.J. Tomkins, The law of the iterated logarithm for weighted sums of independent random variables, J. Theoret. Probab. 16 (2003) 519–542.