Nonparametric quantile estimation under progressive censoring

Sergio Alvarez
Equipe de Mathématiques Appliquées
University of Technology of Compiègne
B.P. 529, 60205 Compiègne - Cedex
France
Sergio.Alvarez@utc.fr

Laurent Bordes
Equipe de Mathématiques Appliquées
University of Technology of Compiègne
B.P. 529, 60205 Compiègne - Cedex
France
Laurent.Bordes@utc.fr

Abstract

This work deals with asymptotic properties of the $[\alpha m]$-th order statistic of a type-II progressively censored sample of size m. Such an order statistic, indexed by $\alpha \in [0,1]$, is called the quantile process. Our main results concern the normalized version of the quantile process for which invariance principles are obtained. These results are applied in order to construct non-parametric estimators of quantiles.

1 Introduction

Order statistics are widely used in statistical modeling and inference. In a paper presenting a unified approach to a variety of model based on order statistics and record values, Kamps (1995) proposed a generalized form of the joint distribution of n ordered random variables. One of the models included in this general setup is the type-II progressive censoring scheme as defined by Balakrishnan and Aggarwala (2000), see also Viveros and Balakrishnan (1994) for an interesting review of the background and of developments in this field. This scheme of censoring has been shown to be of great importance in planning experiments of lifetime data analysis of reliability studies. Hereafter we recall how to obtain such progressively censored data.

Let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) random lifetimes of n items. A progressive type-II right censored sample may be obtained in the following way: at the time of the first failure, noted $X_{1:m:n}$, r_1 surviving items are removed at random from the $n-1$ remaining surviving items, at the time of the next failure, noted $X_{2:m:n}$, r_2 surviving items are removed at random from the $n-r_1-2$ remaining items, and so on. At the time of the m-th failure, all the remaining $r_m = n-m-r_1-\ldots-r_{m-1}$ surviving items are censored. Therefore a progressive type-II right censoring scheme is specified by integers n, m and $r_1, \ldots, r_{m-1}$ with the constraints $n-m-r_1-\ldots-r_{m-1} \geq 0$ and $n \geq m \geq 1$.

If $r_1 = \ldots = r_{m-1} = 0$ we get an usual type-II right censored sample, that is we observe the first m order statistics $X_{1:n}, \ldots, X_{m:n}$ and the $n-m$ remaining times are right censored by $X_{m:n}$. If moreover $m = n$ the usual order statistic is obtained.

Our main developments deals with asymptotic behavior of the $[\alpha m]$-th order statistic of a type-II progressively censored sample of size m ($[x]$ is the unique integer satisfying $[x] \leq x < [x]+1$). Such an order statistic, indexed by $\alpha \in [0,1]$, is called the quantile process and it is written $(X_{[\alpha m]:m:n})_{\alpha \in [0,1]}$. First, we get invariance principles for the normalized version of the quantile process. Therefore, these results are applied in order to construct non-parametric estimators of quantiles (note that recently, by an other approach, Guilbaub (2001) derives exact non-parametric confidence intervals for quantiles). From now on we assume that the X_i's have a common distribution function F with density f. We denote by λ the hazard rate function and Λ the cumulative hazard rate function. The reliability function will be noted $R = 1-F$.

2 Asymptotics for the quantile process

In this section we investigate asymptotic properties of the quantile process by two methods. Let us first introduce the assumptions under which these asymptotic properties are obtained. All asymptotic results will be given with respect to $m \to +\infty$.

A1. $\alpha \in [0, a]$ and $0 < a < 1$;

A2. $(r_i)_{i \geq 1}$ is a bounded sequence of non negative integers ($r_i \leq K$ for all $i \geq 1$);

A3. $\bar{r} = m^{-1} \sum_{i=1}^{m} r_i = r + O(\gamma_m)$ where r is a non negative real number and $(\gamma_m)_{m \geq 1}$ a real valued sequence which satisfies at least $\gamma_m = o(1)$.

A4. There exists a real number $\varepsilon \in [0, a)$ such that λ is continuous and strictly positive on $G^{-1}([\varepsilon, a])$ where $G = 1 - (1 - F)^{r+1}$ is a distribution function.

We now introduce some notations. $(\alpha_j^m)_{1 \leq j \leq m}$ is a triangular array of non negative integers defined by:

$$\alpha_j^m = r_j + \cdots + r_m + m - j + 1 \quad \text{for} \quad 1 \leq j \leq m.$$

We denote by u_α the α-quantile of the distribution function G, $u_\alpha = G^{-1}(\alpha)$ where G^{-1} is taken in the generalized inverse sense ($G^{-1}(x) = \inf\{y \; : \; G(y) > x\}$) if it is not invertible.

2.1 The martingale method

By using a martingale result for partial sums of Dacunha-Castelle and Duflo (1993, Theorem 7.4.28-1, p. 226) we get:

Proposition 1 *Let $\tilde{Y}^{(m)}$ be the process defined on $[0, a]$ by:*

$$\tilde{Y}^{(m)}(\alpha) = m^{1/2} \sum_{j=1}^{[\alpha m]} \frac{Z_j - 1}{\alpha_j^m}.$$

Under A1-A3 we have $\tilde{Y}^{(m)} \xrightarrow{\mathcal{D}} \mathcal{G}_\gamma$ in $D[0, a]$, where $\mathcal{G}_\gamma$ is a Gaussian process on $[0, a]$ with variance function γ defined by:

$$\gamma(\alpha) = \frac{\alpha}{(r+1)^2(1-\alpha)}.$$

Precising the decreasing rate of the sequence $(\gamma_m)_{m \geq 1}$ and using the above proposition, we get:

Corollary 1 *Under A1-A3, if $Y^{(m)}$ is the process defined by*

$$Y^{(m)}(\alpha) = m^{1/2} \left(\sum_{j=1}^{[\alpha m]} \frac{Z_j}{\alpha_j^m} - \frac{1}{r+1} \log\left(\frac{1}{1-\alpha}\right) \right),$$

we have:

(i) $\sup_{\alpha \in [0,a]} \left| m^{-1/2} Y^{(m)}(\alpha) \right| \xrightarrow{P} 0$;

(ii) if $\gamma_m = o(m^{-1/2})$, then $Y^{(m)} \xrightarrow{\mathcal{D}} \mathcal{G}_\gamma$ in $D[0, a]$, where $\mathcal{G}_\gamma$ is the Gaussian process in Proposition 1.

Finally, using the relation between the Y_i's and the X_i's (see Balakrishnan and Aggarwala 2000), the δ-method yields the next convergence result for the quantile process.

Proposition 2 *Under A1-A3 we have:*

(i) $\sup_{\alpha \in [0,a]} \left| X_{[\alpha m]:m:n} - u_\alpha \right| \xrightarrow{P} 0$;

(ii) moreover, if A4 is satisfied, $\gamma_m = o(m^{-1/2})$, $\mathcal{G}_\sigma$ is the Gaussian process on $[\varepsilon, a]$ with variance function σ defined by:

$$\sigma^2(\alpha) = E\left(\mathcal{G}_\sigma^2(\alpha)\right) = \frac{\alpha}{(1+r)^2(1-\alpha)\lambda^2(u_\alpha)}, \quad \alpha \in [\varepsilon, a],$$

and $X^{(m)}$ is the process defined by $X^{(m)}(\alpha) = m^{1/2}(X_{[\alpha m]:m:n} - u_\alpha)$, then, $X^{(m)} \xrightarrow{\mathcal{D}} \mathcal{G}_\sigma$ in $D[\varepsilon, a]$.

2.2 The approximation method

In this section is obtained the asymptotic behavior of the quantile process $(X_{[\alpha m]:m:n})_{\alpha\in[0,a]}$ by using invariance principles. We assume, without loss of generality (Komlòs *et al.* 1976), that all random variables and stochastic processes introduced later on, are defined on the same probability space. We note $\log(m) = \log(m \vee e)$ and $\log_2(m)$ the two iterated logarithm. The proof of the following Law of Iterated Logarithm (L.I.L.), results mainly from the strong law of large numbers for weighted sums of i.i.d. random variables stated in Theorem 1 of Li and Tomkins (2001) and L.I.L. results of Csörgő and Révész (1981, p. 143) for the quantile process of independent and identically distributed random variables.

Proposition 3 *Under conditions A1-A4 with $\gamma_m = o(m^{-1/2})$ we have:*

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

where $\{B_m(t),\ 0 \le t \le 1\}_{m\ge1}$ is a sequence of Brownian bridges. Moreover

$$\lambda(u_\alpha)\sqrt{[\alpha m]}\left(X_{[\alpha m]:m:n} - u_\alpha\right) = \frac{\sqrt{\alpha}}{(r+1)\sqrt{(1-\alpha)}}\frac{W(\alpha m)}{\sqrt{m}} + O_{a.s.}\left(\left(\frac{\log_2 m}{m}\right)^{1/4}(\log m)^{1/2}\right),$$

where $\{W(t)\}_{t\ge0}$ is a standard Wiener process .

3 An estimator of the asymptotic variance

Here we need to estimate $\sigma^2(\alpha) = \alpha/((r+1)^2(1-\alpha)\lambda^2(u_\alpha))$. Since by Proposition 2 and assumption A2, u_α and r are naturally estimated by $X_{[\alpha m]:m:n}$ and $\bar{r} = \sum_{j=1}^m r_j/m$, we propose to estimate σ^2 by:

$$\hat{\sigma}^2(\alpha) = \frac{\alpha}{(\bar{r}+1)^2(1-\alpha)\hat{\lambda}^2(X_{[\alpha m]:m:n})}$$

where $\hat{\lambda}$ is an estimator of λ. Such an estimator, following the ideas of Andersen *et al.* (1993), may be obtained by smoothing the estimator $\hat{\Lambda}^{(n)}$ of the cumulative hazard rate Λ proposed in Bordes (2001). Formally, we define the estimator by:

$$\hat{\lambda}^{(m)}(t) = b_m^{-1}\int_{[0,\tau]} K\left(\frac{t-s}{b_m}\right)\hat{\Lambda}^{(m)}(ds),$$

where $\hat{\Lambda}^{(m)}$ is defined by

$$\hat{\Lambda}^{(m)}(t) = \int_0^t \frac{N^{(m)}(ds)}{Y^{(m)}(s)}.$$

Here processes $N^{(m)}$ and $Y^{(m)}$ are respectively defined by $N^{(m)}(t) = \sum_{j=1}^m 1(X_{j:m:n} \le t)$ and $Y^{(m)}(t) = \sum_{j=1}^m (r_j+1)1(X_{j:m:n} \ge t)$. τ is such that $0 < F(\tau) < 1$, K is a bounded kernel function, which vanishes outside $[-1,1]$ and has integral 1. The bandwidth or window size b_m is a positive parameter satisfying $(b_n)_{n\ge1} \searrow 0$. The martingale approach of Andersen *et al.* (1993) and results in Bordes (2001) give:

Proposition 4 *Under A1–A4 we have $\hat{\sigma}_m^2(\alpha) \xrightarrow{P} \sigma^2(\alpha)$, for all $\alpha \in [\varepsilon, a]$.*

Remark 1 *In Alvarez and Bordes (2002) an other estimator of the variance is proposed. It is based on an approach of Steinebach (1995) which avoids to estimate the hazard rate function.*

4 Reliability example

It is easy to see that if $g(\alpha) = 1 - (1-\alpha)^{r+1}$ then

$$\tilde{q}_\alpha^{(m)} = X_{[g(\alpha)m]:m:n} \xrightarrow{P} F^{-1}(\alpha).$$

Since g is unknown up to the constant r, we define $g^{(m)}(\alpha) = 1 - (1-\alpha)^{\bar{r}+1}$ with $\bar{r} = m^{-1}\sum_{j=1}^m r_j$. Consequently, a natural estimator of the α-quantile of F is

$$\hat{q}_\alpha^{(m)} \equiv X_{[g^{(m)}(\alpha)m]:m:n}.$$

Proposition 5 *Under A1-A4 with $\alpha \in [\varepsilon, a]$, we have:*

$$\hat{q}_\alpha^{(m)} \xrightarrow{P} F^{-1}(\alpha).$$

Moreover, if $\gamma_m = o(m^{-1/2})$, we have:

$$m^{1/2}\left(\hat{q}_\alpha^{(m)} - F^{-1}(\alpha)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(g(\alpha))),$$

where $\sigma^2(g(\alpha))$ can be consistently estimated.

Acknowledgements

Many thanks to Professor Balakrishnan who gave the second author the opportunity to present this work in the MMR'2002 conference.

References

Alvarez, S. and L. Bordes (2002). Quantile process under progressive censoring. Technical report, UTC-EMA.

Andersen, P., O. Borgan, R. Gill, and N. Keiding (1993). *Statistical models based on couting processes.* Springer-Verlag.

Balakrishnan, N. and R. Aggarwala (2000). *Progressive censoring.* Birkhäuser.

Bordes, L. (2001). Non-parametric estimation under progressive censoring. Technical report, UTC-EMA.

Csörgö, M. and P. Révéz (1981). *Strongs approximations in probability and statistics.* Academic Press N. Y.

Dacunha-Castelle, D. and M. Duflo (1993). *Probabilités et statistiques*, Volume 2. Masson. In French.

Guilbaud, O. (2001). Exact non-parametric confidence intervals for quantiles with progressive type-II censoring. *Scand. J. Statist. 28*, 699–713.

Kamps, U. (1995). Non-parametric estimation under progressive censoring. *J. Statist. Plann. Inference 48*, 1–23.

Komlós, J., P. Major, and G. Tusnády (1976). An approximation of partial sums of independent r.v.'s and the sample d.f., part ii. *Z. W. G.. 34*, 33–58.

Li, D. and R. Tomkins (2001). The law of the iterated logarithm for weighted sums of independent random variables. Technical report, University of Regina.

Steinebach, J. (1995). Variance estimation based on invariance principles. *Statist. 27*, 15–25.

Viveros, R. and N. Balakrishnan (1994). Interval estimation of life characteristics from progressively censored data. *Technometrics 36*, 84–91.