Une classe d'estimateurs de type Hodges-Lehmann pour des données censurées

Philippe Capéraà
Département de mathématiques et statistique
Université Laval, Québec, Canada

Abdelaziz Chaoubi
Institut National de Statistique et d'Economie Appliquée (INSEA)
Rabat-Instituts, Rabat, Maroc

SUMMARY

In two papers, Bassiakos, Meng and Lo (1991) and Meng, Bassiakos and Lo (1991) defined a new Hodges-Lehmann type estimator for a shift parameter in the two-sample, random censorship problem. This estimator is based on a truncated version of the Wilcoxon rank statistic, in which the Kaplan-Meier estimators are substituted for the empirical distribution functions. In this paper, the ideas of these authors are extended to a large class of estimators with score functions that are similar to those of Chernoff and Savage (1958). These new estimators are shown to be strongly consistent and asymptotically normal. The bihavior of their asymptotic relative efficiencies is also examined under two stochastic orderings on symmetric distributions. Through simulations, it is seen that the new estimators perform well in the presence of heavy and unequal censoring.

L'estimation d'un paramètre de translation Δ entre deux lois unidimensionnelles continues F et G est un problème qui a été abordé par de nombreux statisticiens, notamment par Hodges et Lehmann (1963) qui ont proposé des estimateurs construits à partir de statistiques de rang. Dans le cas où les deux échantillons provenant des lois F et G sont soumis à un phénomène de censure aléatoire à droite, différentes extensions de ces estimateurs ont été étudiées. Citons entre autres, les estimateurs de Padgett et Wei (1982), de Wei et Gail (1983), de Tsiatis (1990) et de Meng, Bassiakos et Lo (1991). C'est dans ce cadre de données censurées que se situe notre travail.

Le modèle considéré dans ce travail est défini de la façon suivante : soit $X_1, X_2, ..., X_n$ et $Y_1, Y_2, ..., Y_m$ deux échantillons indépendants des lois continues F et G respectivement, telles que $G(t) = F(t-\Delta)$, où Δ est un paramètre inconnu que l'on veut estimer. Soit aussi $U_1, U_2, ..., U_n$ et $V_1, V_2, ..., V_m$ deux autres échantillons, indépendants entre eux et indépendants des deux premiers, provenant des deux lois de censures U et V respectivement. L'objectif est alors d'estimer le paramètre de translation Δ des deux lois F et G après avoir observer $(X_i^c, d_i),\ 1 \le i \le n$ et $(Y_j^c, e_j),\ 1 \le j \le m$, où $X_i^c = \min(X_i, U_i),\ d_i = 1_{(X_i \le U_i)}$, $Y_j^c = \min(Y_j, V_j)$, et $e_j = 1_{(Y_j \le V_j)}$. Soit $\hat{F}_n$ et $\hat{G}_m$ les estimateurs de Kaplan-Meier de F et G, et t_1 et t_2 deux constantes fixées telles que $\hat{F}_n$ et $\hat{G}_m$ convergent presque sûrement vers F et G sur $(-\infty, t_1]$ et $(-\infty, t_2]$ respectivement. Soit J une fonction score définie sur l'intervalle $(0,1)$ et à valeurs dans IR. Considérons alors $(N = n+m)$

$$\hat{P}_{1N} = \int_{-\infty}^{t_1} J\left(\frac{N}{N+1}\hat{F}_n(t)\right) d\hat{F}_n(t), \qquad \hat{P}_{2N} = \int_{-\infty}^{t_2} J\left(\frac{N}{N+1}\hat{G}_m(t)\right) d\hat{G}_m(t)$$

et les deux fonctions :

$$\hat{T}_{1N} = \int_{-\infty}^{t_1} J\left(\frac{n}{N+1}\hat{F}_n(t) + \frac{m}{N+1}\hat{G}_m(t+q)\right) d\hat{F}_n(t), \qquad pour \quad q \le t_2 - t_1$$

et

$$\hat{T}_{2N} = \int_{-\infty}^{t_2} J\left(\frac{n}{N+1}\hat{F}_n(t-q) + \frac{m}{N+1}\hat{G}_m(t)\right) d\hat{G}_m(t), \qquad pour \quad q \geq t_2 - t_1.$$

Soit finalement

$$\hat{\Delta}_{1N}(J) = \min\left\{\hat{T}_{1N}^{-1}(\hat{P}_{1N}), t_2 - t_1\right\} \quad \text{si} \quad \Delta \leq t_2 - t_1$$

et

$$\hat{\Delta}_{2N}(J) = \max\left\{\hat{T}_{2N}^{-1}(\hat{P}_{2N}), t_2 - t_1\right\} \quad \text{si} \quad \Delta \geq t_2 - t_1.$$

L' estimateur $\hat{\Delta}_N(J)$ de Δ dont nous étudions les propriétés asymptotiques est défini par

$$\hat{\Delta}_N(J) = \hat{\Delta}_{1N}(J) + \hat{\Delta}_{2N}(J) - (t_2 - t_1).$$

Remarque. Dans le cas où $J(u) = 2u - 1$ est la fonction score de Wilcoxon, on retrouve l' estimateur étudié par Meng et al. (1991). En outre, lorsque les deux échantillons provenant des deux lois F et G sont entièrement observés (absence de censure), on prend $t_1 = t_2 = +\infty$ et l' estimateur classique de Hodges-Lehmann est égal à . $\hat{T}_{1N}^{-1}(\hat{P}_{1N})$

Sous certaines hypothèses de régularité faites sur la fonctions score J, hypothèses très semblables à celles posées par Chernoff et Savage (1958) dans l' étude des tests linéaires de rang, nous obtenons les résultats suivants :

- $\hat{\Delta}_N(J)$ converge presque sûrement vers Δ ;
- $\hat{\Delta}_N(J)$ a pour loi asymptotique une loi normale.

Le comportement de l' efficacité relative asymptotique de ces nouveaux estimateurs a été aussi examiné avec deux ordres stochastiques pour des distributions symétriques. Enfin, au moyen de simulations, nous montrons que ces estimateurs sont performants en présence de taux de censures élevés et différents.

Références

Bassiakos, Y. G., Meng, X. L. and Lo, S. H. (1991) *A general estimator of the treatment effect when the data are heavly censored*. Biometrika, 78, 741-748.

Capéraà, P. (1988) *Tail ordering and asymptotic efficiency of rank tests. Ann. Statist*., 16, 470-478.

Capéraà, P. and Chaoubi, A. (1995) *A class of Hodges-Lehmann type estimators for censored data*. Technical Report 95-17, Département de mathématiques et de statistique, université Laval, Québec, Canada.

Chernoff, H. and Savage, I. R. (1958) *Asymptotic normality and efficiency of certain nonparametric test statistics*. Ann. Math. Stat., 29, 972-994.

Gill, R. D. (1980) *Censoring and Stochastic Integrals*. Mathematical center Tracts, Vol. 124, Mathematische centrum, Amsterdam.

Hàjek, J. (1969) *A Course in Nonparametric Statistics*. Holden-Day, San Francisco.

Hodges, J. L. and Lehmann, E. L. (1963) *Estimates of location based on rank tests*. Ann. Math. Statist., 34, 598-611.

Loh, W.Y. (1984) *Bounds on ARE's for restrected classes of distributions defined via tail-orderings*. Ann. Statist., 12, 685-701.

Meng, X. L., Bassiakos, Y. G. and LO, S. H. (1991) *Large properties for a general estimator of the treatment effect in the two-sample problem with right censoring*. Ann. Statist., 19, 1786-1812.

Padgett, W. J. and Wei, L. J. (1982) *Estimation of the ratio of scale parameters in the two-sample problem with arbitrary right censorship*. Biometrika, 69, 252-256.

Shorack, G. R. and Wellner, J. A. (1986) *Empirical Processes with Applications to Statistics*. Willey, New York.

Tsiatis, A. A. (1990) Estimating regression parameters using linear rank tests for censored data. Ann. Statist., 18, 354-372.

Wei, L. J. and Gail, M. H. (1983) *Nonparametric estimation for scale-change with censored observation*, J. Amer. Statisti. Assoc., 78, 382-388.