Serdica

Bulgariacae mathematicae publicationes

Сердика

Българско математическо списание

SOME STATISTICAL TESTS ASSOCIATED WITH THE CONCEPT OF Δ-STOCHASTIC ORDERING OF TWO RANDOM VARIABLES

R. E. GOOT, S. T. RACHEV, A. YU. YAKOVLEV, N. O. KADYROVA, G. M. ZHARINOV

The present paper deals with various generalizations of some known partial ordering relations for probability distributions. The usual inequalities are weakened by means of an additive term Δ. Some tests for the hypothesis about these Δ-stochastic ordering relations are proposed. Furthermore, the problem of estimating a lower confidence limit for the parameter Δ is considered. An example with survival data illustrates the use of the proposed methods.

1. Introduction. Various statistical tests for comparing the cumulative distributions in survival data analysis have been fomulated in the statistical literature by many authors, e. g. [1, 3, 5, 9]. However they have paid less attention to the stochastic ordering of random variables. A new notion of Δ-stochastic ordering of two random variables has been introduced in [7] and it has proved to be of some use for the survival analysis. In this paper some approaches to the statistical testing of Δ-stochastic ordering relations are developed.

The random variable T_2 is said to be Δ-stochastically smaller than the variable T_1 written as $T_2 \underset{\Delta-\mathrm{st}}{\overset{(1)}{<}} T_1$, if for a given $\Delta \geqq 0$ and every $t \in \mathrm{R}^1$

$$F_1(t)-\Delta \leqq F_2(t), \tag{1.1}$$

where F_1 and F_2 are the distribution functions of T_1 and T_2, respectively. For $\Delta=0$ relation (1.1) coincides with the well-known definition (see [8]) of the ordering relation $\overset{(1)}{\leqq}$ for random variables. The effect of the above generalization of the relation $\overset{(1)}{\leqq}$ for the statistical inference can be illustrated by the following example.

Let $\{F_{X_n}\}$ be a sequence of distribution functions of the random variables $\{X_n\}$ and let $\{F_{Y_n}\}$ be a similar sequence for the random variables $\{Y_n\}$. The uniform convergence, i. e. convergence in the Kolmogorov metric ρ,

$$\rho(F_X, F_Y)=\sup\{|F_X(t)-F_Y(t)|: t \in \mathrm{R}^1\},$$

of $F_{X_n}(t)$ and $F_{Y_n}(t)$, respectively, to $F_X(t)$ and $F_Y(t)$, where $F_X \leqq F_Y$, does not imply that there exists n_0 such that $F_{X_n} \leqq F_{Y_n}$ is true for all $n > n_0$. However, in this case it may be stated that $Y_n \underset{\Delta_n-\mathrm{st}}{\overset{(1)}{<}} X_n$, if Δ_n is chosen as follows:

$$\Delta_n=\rho(F_{X_n}, F_X)+\rho(F_{Y_n}, F_Y).$$

The present paper deals with statistical tests for hypotheses concerning that type of relations between the random variables.

2. Statistical tests for Δ-stochastic ordering relations. Having in mind the applications to survival data analysis we consider the nonnegative random variables T_1, T_2 and assume $F_1(0)=F_2(0)=0$. Further the same notations of ordering relations as those in [8] will be used. In the context of the Δ-stochastic ordering concept two problems can be examined:

SERDICA Bulgaricae mathematicae publicationes. Vol. 16, 1990, p. 240—245.

a) For some specified value $\Delta=\Delta^*\geqq 0$ in the relation $T_2 \underset{\Delta-st}{\overset{(1)}{<}} T_1$ test the hypothesis H_0: $\Delta=\Delta^*$ (that is for all t, $F_1(t)-\Delta^*\leqq F_2(t)$) against the alternative $H_1:\Delta>\Delta^*$;
b) Estimate the lower confidence limit $\Delta_{\min}$ for $\Delta\geqq 0$ from the ordering relation $T_2 \underset{\Delta-st}{\overset{(1)}{<}} T_1$.
Consider problem (a).

Let $F_1^{(n)}$ and $F_2^{(k)}$ be the empirical distribution functions corresponding to F_1 and F_2. Let us denote by n and k the sample sizes for the random variables T_1 and T_2, respectively.

If the hypothesis under test holds true, then

$$F_1^{(n)}(t)-F_2^{(k)}(t)\leqq\Delta^*+\rho(F_1^{(n)}, F_1)+\rho(F_2^{(k)}, F_2),\quad t\geqq 0,$$

where $\rho(F^{(\cdot)}, F)$ is the one-sample Kolmogorov — Smirnov statistic. Consequently, for independent samples and any $c>0$

$$P\{F_1^{(n)}(t)-F_2^{(k)}(t)\leqq\Delta^*+c\}\geqq \sup_{0<h<c}\ [G^{(n)}(h)G^{(k)}(c-h)],$$

where $G^{(m)}$ is the Kolmogorov distribution for a sample of size m. The constant c is fixed or may also be chosen to maximize the right-hand side.

The statistical test obtained for the hypothesis H_0: $T_2 \underset{\Delta-st}{\overset{(1)}{<}} T_1$ is determined under the significance level

$$\alpha\leqq 1-\sup_{0<h<c}\ [G^{(n)}(h)G^{(k)}(c-h)], \tag{2.1}$$

with the critical domain in the sample space

$$F_1^{(n)}(t)-F_2^{(k)}(t)>\Delta^*+c,\quad t\geqq 0.$$

It is not necessary to require sample independence, if instead of (2.1), bounds for the joint distributions with fixed marginal distributions [2, 6] are used. Then we have

$$\alpha\leqq 1-\max\{\sup_{0<h<c}\ [G^{(n)}(h)+G^{(k)}(c-h)-1],\ 0\}. \tag{2.2}$$

Similarly the ordering relation $\overset{(L)}{<}$ and an appropriate test can be considered. In this case we define Δ-stochastic ordering $T_2 \underset{\Delta-st}{\overset{(L)}{<}} T_1$ as follows:

$$u\int_0^\infty e^{-ut}F_1(t)dt-\Delta\leqq u\int_0^\infty e^{-ut}F_2(t)dt,\quad u>0.$$

The test rule for the hypothesis H_0 will be of the form

$$\sup_{u>0}\{\frac{1}{n}\sum_{i=1}^{n} e^{-uT_1^{(i)}}-\frac{1}{k}\sum_{i=1}^{k} e^{-uT_2^{(i)}}\}\leqq\Delta+c,$$

where $T_j^{(i)}$, $j=1, 2$, are the i-th order statistics obtained from j-th sample. The significance level is also calculated from (2.1) or (2.2).

In a more general case Δ-stochastic ordering $T_2 \underset{\Delta-st}{\overset{(N)}{<}} T_1$ may be introduced as

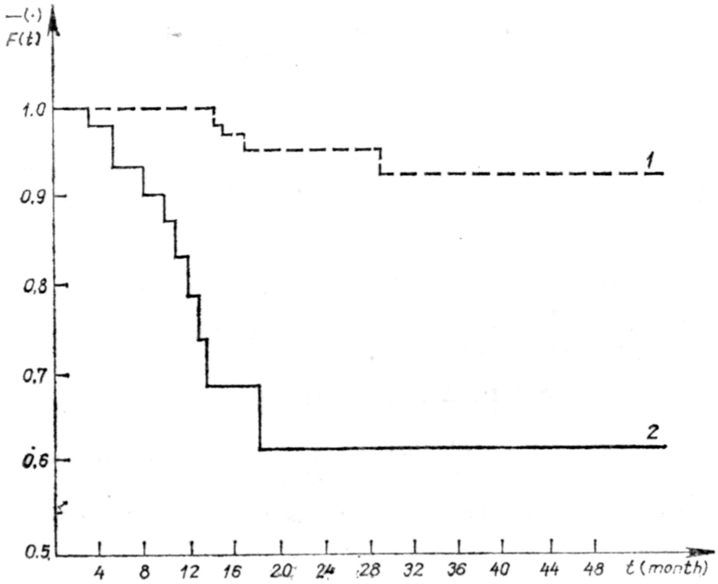

Fig. 1. The product-limit estimates of survival functions for two groups of oncological patients

follows

$$\int_0^\infty \varphi(t, u) F_1(t)\, dt - \Delta \leq \int_0^\infty \varphi(t, u) F_2(t)\, dt. \tag{2.3}$$

The relation (2.3) is defined for all nonnegative functions φ providing the existence of integrals on both sides of the inequality. In particular, if $\varphi(t, u) = \delta(t-u)$, where $\delta(\cdot)$ is the Dirac delta-function, we formally have the relation $\underset{\Delta-\mathrm{st}}{\overset{(1)}{<}}$; $\varphi(t, u) = ue^{-ut}$ corresponds to the relation $\underset{\Delta-\mathrm{st}}{\overset{(L)}{<}}$. For $\varphi(t, u) = \Theta(u-t)$, where $\Theta(x) = 0$ for $x < 0$ and $\Theta(x) = 1$ for $x \geq 0$, the relation $\underset{\Delta-\mathrm{st}}{\overset{(3)}{<}}$ is valid (relation $\underset{\Delta-\mathrm{st}}{\overset{(3)}{<}}$ is the Δ-stochastic analogue of relation $\overset{(3)}{\leq}$, (see [8]). Also suppose that the functions $\varphi(t, u)$ satisfy the condition

$$0 < \int_0^\infty \varphi(t, u)\, dt = v(u) < \infty, \quad u \in [0, \infty).$$

Let the hypothesis H_0: $T_2 \underset{\Delta-st}{\overset{(N)}{<}} T_1$ hold true; i. e. (2.3) is valid. Then the following inequality holds:

$$\int_0^\infty \varphi(t, u) [F_1^{(n)}(t) - F_2^{(k)}(t)]\, dt \leq \Delta + v(u) [\rho(F_1^{(n)}, F_1) + \rho(F_2^{(k)}, F_2)].$$

Hence let us put $V=\sup_{u\geq 0} v(u)$. Finally we obtain a statistical test for H_0. The null hypothesis will be accepted if under given sample observations we have

$$\frac{1}{n}\sum_{i=1}^{n}\varphi(T_1^{(i)},u)-\frac{1}{k}\sum_{i=1}^{k}\varphi(T_2^{(i)},u)\leq\Delta+\frac{c}{V},$$

otherwise it should be rejected.

Next the significance level is given by

$$\alpha\leq 1-\max\{\sup_{0<h<c}[G^{(n)}(\frac{h}{V})+G^{(k)}(\frac{c-h}{V})-1],\ 0\}.$$

3. A test based on a preordering relation between the distribution functions. Let us consider another statistical test which seems to be more natural for comparing the survival functions of two independent populations. We use (2.3) again, setting $\varphi(t,u)dt=dF_2(t)$, $\Delta=0$. Then we have

(3.1) $$\int_0^\infty F_1(t)\,dF_2(t)\leq 1/2.$$

It is evident that (3.1) defines only the preordering relation with respect to the distributions F_1 and F_2, because the antisymmetry condition is not fulfilled here.

The Δ-stohastic version of (3.1) takes the form

(3.2) $$\int_0^\infty F_1(t)dF_2(t)\leq 1/2+\Delta,$$

and if inequalitly (3.2) holds, then $T_2 \overset{(G)}{\underset{\Delta-\mathrm{st}}{<}} T_1$.

If T_1 and T_2 are independent random variables, then (3.2) may be rewritten in the following form

$$P\{T_2>T_1\}-2\Delta\leq P\{T_1>T_2\},$$

which explicitly explains the reason why the relation $\overset{(G)}{\underset{\Delta-\mathrm{st}}{<}}$ is useful in survival data analysis.

It is easy to show that

$$\int_0^\infty F_1^{(n)}(t)dF_2^{(k)}(t)\leq\rho(F_1^{(n)},F_1)-\rho(F_2^{(k)},F_2)+\int_0^\infty F_1(t)dF_2(t).$$

Assume that the hypothesis under test (inequality (3.2)) holds. Then we have

$$\int_0^\infty F_1^{(n)}(t)dF_2^{(k)}(t)\leq 1/2+\Delta+\rho(F_1^{(n)},F_1)+\rho(F_2^{(k)},F_2).$$

Hence the test rule for the hypothesis H_0: $T_2 \overset{(G)}{\underset{\Delta-\mathrm{st}}{<}} T_1$ will be of the form

$$\int_0^\infty F_1^{(n)}(t)dF_2^{(k)}(t)\leq 1/2+\Delta+c.$$

Thus the significance level is calculated by analogy with (2.1) or (2.2).

4. Lower confidence limit for the parameter Δ. The solution of Problem (b) for the above-mentioned Δ-stochastic ordering relations may be obtained as follows. Let

$$\Delta(n,\ k)=\min\{\Delta\geqq 0:\ F_1^{(n)}(t)-F_2^{(k)}(t)\leqq\Delta\}.$$

Then we have

$$F_1(t)-F_2(t)\leqq\Delta(n,\ k)+\rho(F_1^{(n)},\ F_1)+\rho(F_2^{(k)},\ F_2), \tag{4.1}$$

Hence, for any $c>0$

$$\begin{aligned} P\{\Delta\geqq\Delta(n,\ k)+c\} &\geqq \sup_{0<h<c} P(\{\rho(F_1^{(n)},\ F_1)<h,\ \rho(F_2^{(k)},F_2)<c-h\} \\ &\geqq \max\{\sup_{0<h<c}[G^{(n)}(h)+G^{(k)}(c-h)-1],\ 0\}. \end{aligned} \tag{4.2}$$

Thus $\Delta_{\min}=\Delta(n,\ k)+c$ is an estimate of the minimal Δ-value for which the ordering relation $\underset{\Delta-\mathrm{st}}{\overset{(1)}{<}}$ is attained with probability greater than the value on the right side of (4.2). Notice that $c=c(n,\ k,\ \alpha)$.

If the relation $\underset{\Delta-\mathrm{st}}{\overset{(L)}{<}}$ holds, let

$$\Delta(n,\ k)=\min\{\Delta\geqq 0:\ u\int_0^\infty F_1^{(n)}(t)\,e^{-ut}\,dt-u\int_0^\infty F_2^{(k)}(t)\,e^{-ut}\,dt\leqq\Delta\}.$$

Then it is easy to show that

$$u\int_0^\infty e^{-ut}F_1(t)\,dt-u\int_0^\infty e^{-ut}F_2(t)dt\leqq\Delta(n,\ k)+\rho(F_1^{(n)},\ F_1)+\rho(F_2^{(k)},\ F_2).$$

Consequently, $\Delta_{\min}=\Delta(n,\ k)+c$ for the relation $\underset{\Delta-\mathrm{st}}{\overset{(L)}{<}}$ by analogy with (4.2) gives the respective probability.

Consider the preordering relation $\underset{\Delta-\mathrm{st}}{\overset{(G)}{<}}$. Let

$$\Delta(n,\ k)=\min\{\Delta\geqq 0:\ \int_0^\infty F_1^{(n)}(t)\,dF_2^{(k)}(t)-1/2\leqq\Delta\}.$$

It can be shown that

$$\int_0^\infty F_1(t)dF_2(t)-1/2\leqq\Delta(n,\ k)+\rho(F_1^{(n)},\ F_1)+\rho(F_2^{(k)},\ F_2).$$

Then, by analogy with (4.2), $\Delta_{\min}=\Delta(n,\ k)+c$ is the lower confidence limit of the parameter Δ for the relation $\underset{\Delta-\mathrm{st}}{\overset{(G)}{<}}$.

Using the results by Hall and Wellner [4] it is possible to obtain some modifications of the above tests for censored (within scheme of right-hand independent censorship) observations in large sample studies. In that case one might replace the usual empirical distributions by the product-limit estimators for the survival functions and Kolmogorov distribution — by the asymptotic sample distribution tabulated in [4].

5. Numerical example. The patterns of the empirical counterparts $\bar{F}_1^{(n)}(t)$, $\bar{F}_2^{(k)}(t)$ of the survival functions $\bar{F}_1=1-F_1$ and $\bar{F}_2=1-F_2$, respectively, are shown in Fig. 1. The functions $\bar{F}_1^{(n)}$ and $\bar{F}_2^{(k)}$ have been computed by using real survival data due to observing oncological patients after radiation therapy. Two groups of patients with carcinoma of uterine cervix have been observed: these groups differed from each other as regards to the disease stage. The first group comprises $n=146$ patients (stage 2) and the second one — $k=56$ patients (stage 3). Both samples are subject to random right censoring.

The following estimates of Δ_{min}, i. e. problem (b), have been obtained for $P\{\Delta \geqq \Delta_{min}\}=0.95$:

1) for the ordering relation $T_2 \overset{(\cdot)}{\underset{\Delta-st}{<}} T_1$:

$$\Delta_{min}=0.672 \text{ for type (1): } (\Delta(n, k)=0.331)$$

$$\Delta_{min}=0.552 \text{ for type } (L); \ (\Delta(n, k)=0.181)$$

$$\Delta_{min}=0.666 \text{ for type } (G); \ (\Delta(n, k)=0.325)$$

2) for the ordering relation $T_1 \overset{(\cdot)}{\underset{\Delta-st}{<}} T_2$:

$$\Delta_{min}=0.341 \text{ for type (1); } (\Delta(n, k)=0)$$

$$\Delta_{min}=0.341 \text{ for type } (L); \ (\Delta(n, k)=0)$$

$$\Delta_{min}=0.341 \text{ for type } (G); \ (\Delta(n, k)=0).$$

REFERENCES

1. I. Bagai, J. V. Deshpandé, S. C. Kochar. Distribution free tests for stochastic ordering in the competing risks model. Tech. Rep. No 37, 1987.
2. M. Fréchet. Sur les tableaux de correlation dont les marquees sont données. *Ann. Univ. Lyon.*, Sect. A, Ser. 3, **14**, 1951, 53-77.
3. R. Gray. A class of k-sample tests for comparing the cumulative incidence of a competing risk. *Ann. Statist.*, **16**, 1988, No 3, 1141-1151.
4. W. J. Hall, J. A. Wellner. Confidence bands for a survival curve from censored data. *Biometrika*, **67**, 1980, 133-143.
5. D. P. Harrington, T. R. Fleming. A class of rank test procedures for censored survival data. *Biometrika*, **69**, 1982, No 3, 553-566.
6. W. Höeffding. Masstabsinvariante Korrelationstheorie. *Schr. Math. Inst. Univ.* Berlin, **5**, 1940, 181-233.
7. S. T. Rachev, A. Yu. Yakovlev, N. O. Kadyrova, E. M. Myasnikova. On the statistical inference from survival experiments with two types of failure. *Biometrical J.*, **30**, 1988, 835-842.
8. D. Stoyan. Qualitative Eigenschaften und Abschat-Zungen Stochastischer Modelle. Berlin, 1977.
9. H. K. Ury. On distribution-free confidence bounds for $P\{X<Y\}$. *Technometrics*, **11**, 1972, 577-582.

Prof. A. Yu. Yakovlev
Department of Applied Mathematics
Leningrad Polytechnical Institute
Polytechnicheskaya ul., 29
Leningrad 195251, USSR

Received 16. 06. 1989
Revised 30. 03. 1990