Ann. Inst. Statist. Math.
37 (1985), Part A, 529-533

APPROXIMATIONS TO THE DISTRIBUTIONS OF ORDERED DISTANCE RANDOM VARIABLES

R.-D. REISS

(Received Mar. 12, 1984; revised Nov. 12, 1984)

Summary

In the present note we give short proofs of asymptotic theorems for the distributions of extreme and intermediate ordered distance random variables. Moreover, a quick goodness-of-fit test is proposed which is based on a single intermediate ordered distance random variable.

1. Introduction

Let $X_0, X_1, X_2, \cdots$ be independent random vectors with values in the m-dimensional Euclidean space R^m. Denote by P_0 the distribution of X_0. Moreover, let $X_1, X_2, \cdots$ have a common Lebesgue-density f_0. It is clear that for every norm $\| \ \|$ on R^m and every fixed point $x_0 \in R^m$ the random distances $\|X_i - x_0\|$, $i=1, 2, \cdots$, are independent and identically distributed with the common distribution function $F(x_0, r)=\int_{B(x_0,r)} f_0(x)dx$ for $r \geqq 0$ where $B(x_0, r)=\{y : \|y-x_0\| \leqq r\}$ is the ball with center x_0 and radius r. Hereafter, $V(r)$ denotes the volume of $B(x_0, r)$. We have $V(r)=K_m r^m$ where e.g. $K_m=\pi^{n/2}/\Gamma(n/2+1)$ in the particular case of the Euclidean distance with Γ denoting the gamma function. Let $Z_{k:n}: R^n \rightarrow R$ be defined by $Z_{k:n}(x)=z_k$ where $z_1 \leqq \cdots \leqq z_n$ are the components of $x=(x_1, \cdots, x_n)$ in the nondecreasing order. Then the kth nearest neighbour distance is defined by $Z_{k:n}((\|X_i - x_0\|)_{i=1}^n)$. More generally, we define $R_{k:n}=Z_{k:n}((\|X_i - X_0\|)_{i=1}^n)$ as the kth ordered distance random variable. The asymptotic distribution of $R_{k:n}$ has been studied by several authors (see e.g. Dziubziella [1] and Mammitzsch [6] and the references given there). The proof of Dziubdziella is based on a limit theorem for exchangeable random variables. Mammitzsch was able to give a different proof without a boundedness condition on f which was needed by Dziubdziella. We shall generalize the result of Mammitzsch

Key words: Ordered distance random variables, approximate distribution, gamma distributions.

which is given here as Corollary 2.2 and shorten the proof considerably.

2. Asymptotic theorems

Our main result will be proved under the following two conditions on $k\equiv k(n)$, f_0, P_0 and certain standardizing functions $r(\cdot, n)$:

CONDITION A. $k(n)/n\to 0$ as $n\to\infty$.

CONDITION B. For P_0 almost all x_0:

$$F(x_0, r(x_0, n)n^{-1/m})=V(r(x_0, n)n^{-1/m})\cdot f_0(x_0)+o(k(n)^{1/2}/n)$$

as $n\to\infty$, and $r(\cdot, n)$ is measurable.

If $k(1), k(2),\cdots$ is a bounded sequence then Condition B is fulfilled for every f_0 being continuous. Finally, let L_k denote the distribution function of the gamma distribution with parameter k; that is, $L_k(x)=1-\exp(-x)\sum_{i=0}^{k-1} x^i/i!$ for $x\geqq 0$.

THEOREM 2.1. *If the Conditions A and B are fulfilled then*

$$P\{n^{1/m}R_{k(n):n}\leqq r(X_0, n)\}=\int L_{k(n)}(V(r(\cdot, n))f_0)dP_0+o(1) ,$$

PROOF. Fubini's theorem implies that

$$P\{n^{1/m}R_{k(n):n}\leqq r(X_0, n)\}=\int P\{n^{1/m}Z_{k(n):n}((\|X_i-x_0\|)_{i=1}^n)\leqq r(x_0, n)\}dP_0(x_0) .$$

Thus, in view of Lebesgue's dominated convergence theorem it remains to prove that

$$P\{n^{1/m}Z_{k(n):n}((\|X_i-x_0\|)_{i=1}^n)\leqq r(x_0, n)\}=L_{k(n)}(V(r(x_0, n))f_0(x_0))+o(1)$$

for those x_0 which fulfill Condition B. By [7], Theorem 2.1, or [8], Theorem 2.6, and by the probability integral transformation and simple straightforward calculations we obtain

$$\begin{aligned}P\{n^{1/m}&Z_{k(n):n}((\|X_i-x_0\|)_{i=1}^n)\leqq r(x_0, n)\}\\ &=L_{k(n)}(nF(x_0, r(x_0, n)/n^{1/m}))+o(1)\\ &=L_{k(n)}(V(r(x_0, n))f_0(x_0)+o(k(n)^{1/2}))+o(1)\\ &=L_{k(n)}(V(r(x_0, n))f_0(x_0))+o(1) .\end{aligned}$$

The easiest way to prove the last relation is to apply the inequality $|L_k(t)-\Phi((t-k)/k^{1/2})|\leqq Ck^{-1/2}$ (see [8], page 540) where $C>0$ is a universal constant and Φ denotes the standard normal distribution function. The proof is complete.

From Theorem 2.1 we derive the Theorem in Mammitzsch [6] as a special case.

COROLLARY 2.2. *If r and k are fixed and P_0 is absolutely continuous with respect to the Lebesgue-measure then*

$$P\{n^{1/m}R_{k:n}\leqq r\}=\int L_k(V(r)f_0)dP_0+o(1)\ .$$

PROOF. Condition A is trivially fulfilled. Moreover Condition B can easily be verified by making use of the fact that Lebesgue a.a. x_0 are Lebesgue-points of f_0 (see e.g. [2], page 276).

When proving Corollary 2.2 in a direct way one can apply Theorem 2.8.2 in [3]—with $\gamma=m$ and $d_n=(K_m f_0(x_0)n)^{-1/m}$—in place of the results in [6] and [8].

Our Theorem 2.1 generalizes the result in [6] with respect to the following two aspects: Firstly, $k(n)$ is not necessarily fixed but may also go to infinity as $n\to\infty$ and, secondly, since it is not supposed in Theorem 2.1 that P_0 is absolutely continuous we include in our considerations e.g. the case of the nearest neighbour distance.

Under the Conditions A and B we obtain at once from [8], page 540, that

$$P\{n^{1/m}R_{k(n):n}\leqq r(X_0, n)\}=\int \Phi((V(r(\cdot, n))f_0-k(n))/k(n)^{1/2})dP_0+o(1)\ .$$

If r is fixed then it is clear that the result is degenerate in the sense that the right-hand side converges to zero. If $r(x_0, n)=[(k(n)+tk(n)^{1/2})/(K_m f_0(x_0))]^{1/m}$, then

$$P\{n^{1/m}R_{k(n):n}\leqq r(X_0, n)\}=\Phi(t)+o(1)\ .$$

We remark that Condition B can e.g. be verified under Lipschitz or differentiability conditions on f_0 (see also Section 3). In the special case of P_0 being a Dirac-measure at x_0 one obtains an asymptotic theorem for the $k(n)$th nearest neighbour distance. In the case that f_0 has second partial derivatives such a result is indicated in [4].

3. Extensions and a quick goodness-of-fit test

Throughout this section let P_0 be a probability measure with Lebesgue-density f_0, and h is a uniformly bounded function with $\int hdP_0=0$. Moreover, $X_0,\cdots, X_n$ are iid random vectors which have a common Lebesgue-density $f_n=f_0(1+h/k(n)^{1/2})$. Thus, for $h\equiv 0$ we have the situation of Section 2 in the case that $X_0,\cdots, X_n$ are identically distributed with distribution P_0. Denote by $F(x_0, n, \cdot)$ the distribution function of $\|X_0 - x_0\|$; that is, $F(x_0, n, r) = \int_{B(x_0, r)} f_n(x)dx$.

CONDITION B′. For P_0 almost all x_0:

$$F(x_0, n, r(x_0, n)n^{-1/m}) = V(r(x_0, n)n^{-1/m})f_n(x_0) + o(k(n)^{1/2}/n)$$

as $n \to \infty$, and $r(\cdot, n)$ is measurable.

Thus Condition B′ reduces to Condition B if $h=0$. Using the arguments of Section 2 we obtain

THEOREM 3.1. *If the Conditions A and B′ are fulfilled then*

$$P\{n^{1/m}R_{k(n):n} \leqq r(X_0, n)\} = \int L_{k(n)}(V(r(\cdot, n))f_n)dP_0 + o(1) .$$

For testing the null-hypothesis f_0 define the critical region

$$C_{\alpha,n} = \left\{k(n)^{1/2}f_0(X_0)\left|\frac{K_m n}{k(n)}R^m_{k(n):n} - \frac{1}{f_0(X_0)}\right| > c(\alpha)\right\}$$

where $\alpha \in (0, 1)$ and $c(\alpha) = \Phi^{-1}(1-\alpha/2)$. Roughly speaking, the critical region is defined by means of a density estimator which is evaluated at a random point. It is straightforward to conclude from Theorem 3.1 that

$$P(C_{\alpha,n}) = 2\left[1 - \int \frac{\Phi(c(\alpha)+h) + \Phi(c(\alpha)-h)}{2} dP_0\right] + o(1) \tag{3.2}$$

if $k(n) \to \infty$ as $n \to \infty$ and Condition B′ holds for $r(x_0, n) = [(k(n) \pm c(\alpha) \cdot k(n)^{1/2})/(K_m f_0(x_0))]^{1/m}$.

Under the null-hypothesis (that is, $h=0$) we have

$$P(C_{\alpha,n}) = \alpha + o(1) . \tag{3.3}$$

If $P_0\{h \neq 0\} > 0$ then the right-hand side of (3.2) is asymptotically equal to some number $\beta \in (\alpha, 1)$. This can be proved by using e.g. Anderson's lemma.

Assume that the support $T = \{x: f_0(x) > 0\}$ of f_0 is open and that f_0 and h have bounded second partial derivatives on T. Then Condition B′ holds true for every sequence $k(n) = o(n^{4/(m+4)})$ and every h which is uniformly bounded and has the property $\int h dP_0 = 0$. In [5] a sequence of tests was defined which is based on variables of the type $R_{1:n}$. These tests asymptotically attain a rejection probability $\beta \in (\alpha, 1)$ under contiguous alternatives of the form $f_0(1+h/\log n)$ if f_0 and h fulfill certain weak smoothness conditions. Our results show that quick tests of a better performance can be found if the smoothness conditions on f_0 and h are slightly strengthened.

Acknowledgement

Prof. Mammitzsch kindly provided us with a preprint of his article.

UNIVERSITY OF SIEGEN

REFERENCES

[1] Dziubziella, W. (1976). A note on the kth distance random variables, *Zastosowania Matematykai*, **15**, 289-291.

[2] Floret, K. (1981). Maß- und Integrationstheorie, *Teubner*, Stuttgart.

[3] Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics*, Wiley, New York.

[4] Hall, P. (1983). On near neighbour estimates of a multivariate density, *J. Mult. Analysis*, **13**, 24-39.

[5] Henze, N. (1982). The limit distribution for maxima of 'weighted' rth-nearest-neighbourhood distances, *J. Appl. Prob.*, **19**, 344-354.

[6] Mammitzsch, V. (1983). On the kth ordered distance random variables, *Proceedings of the Third Prague Conference on Asymptotic Statistics* (to appear).

[7] Matsunawa, T. and Ikeda, S. (1976). Uniform asymptotic distributions of extremes. In: Essays in Probability and Statistics (ed. Ikeda et al), Shinko Tsushi Tokyo, 419-432.

[8] Reiss, R.-D. (1981). Uniform approximation to distributions of extreme order statistics, *Adv. Appl. Prob.*, **13**, 533-547.