The Annals of Applied Probability
1993, Vol. 3, No. 2, 454–466

ASYMPTOTIC DISTRIBUTION OF THE NORMAL SAMPLE RANGE

BY PETER C. MATTHEWS[1] AND ANDREW L. RUKHIN[2]

University of Maryland

For a spherically symmetric multivariate normal random sample, the asymptotic distribution of the largest interpoint Euclidean distance is derived. The number of interpoint distances exceeding a high level is shown to have a limiting Poisson distribution.

1. Introduction and summary. In this paper the asymptotic distribution of the multivariate normal range is determined. Let $X_1, X_2, \dots$ be independent k-dimensional normal random vectors with zero mean vector and the identity covariance matrix. The sample range M is defined as the largest interpoint distance between the first n observations:

$$M = M(n) = \max_{1 \le i < j \le n} |X_i - X_j|. \tag{1.1}$$

The asymptotic distribution of M is well known in the special case $k = 1$ [see David (1981), Section 9.4]; namely, for any c,

$$\begin{aligned}
&\lim_{n \mapsto \infty} P\big(M^2 \le 4[2\log n - \log\log n - \log 4\pi + c]\big)\\
&\quad = \lim_{n\to\infty} P\left(\sqrt{2\log n}\left[M - 2\sqrt{2\log n} + \frac{\log\log n + \log 4\pi}{2\sqrt{2\log n}}\right] \le c\right)\\
&\quad = \int_{-\infty}^{\infty} \exp(-c - e^{-t-c} - e^{-t})\, dt;
\end{aligned}$$

that is, the asymptotic distribution of the range is the convolution of the limiting distributions for the extreme order statistics. The situation when $k \ge 2$ is considerably less studied, although the problem is interesting both from theoretical and practical points of view.

Indeed for $k = 2$ the range is an important statistic in gun quality control, in which a gun is accepted when its sample range of points of impact does not exceed a given value. For instance, in the quality control program of Smith & Wesson Co., a handgun is placed in a vise and fired 10 times at a target with a grid on it, so that the determination of the largest distance between points

Received August 1991; revised May 1992.
[1] Research supported by NSF and NSA under NSF Grant 90-01295
[2] Research supported by NSF Grant 90-00999.
AMS 1991 *subject classifications*. Primary 60E05; secondary: 60F99, 62E20, 62H10.
Key words and phrases. Normal sample range, U-statistic, Poisson clumping heuristic, exceedance.

of impact is an easy task. If this distance exceeds 4 inches, the gun is rejected. In this situation the calculation of the sum of squares needed for the optimal test for the dispersion of points of impact is time consuming and not always feasible. We will discuss the distribution theory relevant to this example in Section 4. The monograph of Grubbs (1964) discusses further statistical measures of accuracy in bivariate samples.

In general situations the range could be used for detecting outliers [cf. Barnett and Lewis (1984), Chapter 9.3] or to provide a quick, short-cut estimate of the dispersion [cf. David (1981), Section 7.5, where this statistic is described as "intriguing"].

S. Wilks obtained by the Monte Carlo method the first four moments of M for some values of sample size n, which are reproduced in Cacoullos and DeCiccio (1967). These authors also suggested a chi approximation to the distribution of M and noticed that a lower bound to its distribution function can be obtained from the distribution of the diameter of the smallest circle covering the whole sample. The latter was derived in Daniels (1952). Siotani (1959) discussed some bounds for the percentiles of the distribution of M.

The asymptotic distribution of the smallest interpoint distance was studied by Silverman and Brown (1978) and Jammalamadaka and Janson (1986) by using the Poisson limit theorem for U-statistics [see also Barbour and Eagleson (1984)]. A related result on the asymptotic behavior of random points with specified nearest neighbor relations was obtained by Henze (1987).

Here is the formulation of our main result for $k \geq 2$. Let

$$l_2 n = \log\log n \quad \text{and} \quad l_3 n = \log l_2 n.$$

Throughout this paper we assume that n is sufficiently large and that natural logarithms are used. Denote for fixed c,

$$r_1 = \left[2\log n + \tfrac{1}{2}(k-3)l_2 n + l_3 n + a + c\right]^{1/2} \tag{1.2}$$

with

$$a = a(k) = \log\frac{(k-1)2^{(k-7)/2}}{\Gamma(k/2)\pi^{1/2}}. \tag{1.3}$$

Let

$$C_n = \operatorname{card}\{(i,j): 1 \leq i < j \leq n, |X_i - X_j| \geq 2r_1\}$$

be the number of exceedances by the interpoint distances of the given level $2r_1$.

THEOREM 1. *As n tends to infinity, C_n converges in distribution to a Poisson random variable with parameter e^{-c}.*

This theorem has the following corollary, which is a consequence of the equivalence of the events $C_n > 0$ and $M > 2r_1$.

COROLLARY 1. *For $k \geq 2$,*

$$\lim_{n\to\infty} P\left(M^2 \leq 4\left[2\log n + \frac{1}{2}(k-3)l_2 n + l_3 n + a + c\right]\right)$$
$$= \lim_{n\to\infty} P\left(\sqrt{2\log n}\left[M - 2\sqrt{2\log n} - \frac{0.5(k-3)l_2 n + l_3 n + a}{2\sqrt{2\log n}}\right] \leq c\right)$$
$$= \exp(-e^{-c}).$$

In this paper, the term "extreme value distribution" will refer to the cumulative distribution function $\exp(-e^{-c})$, possibly with location and scale changes. The first expression, based on M^2, is included because the distribution of M^2 in simulations is closer than the distribution of M is to an extreme value distribution.

The exceedance count C_n is related to the Poisson clumping heuristic defined in Aldous (1989). Intuitively, a vector X_i with an exceptionally large norm could lead to a clump of exceedances of $2r_1$ when combined with vectors of large norm and nearly opposite direction. Indeed it looks as if this possibility prevents the moments of C_n from converging to the moments of the limiting Poisson distribution. For example, when $k = 2$,

$$EC_n = \binom{n}{2} P\left(Y > 2\left[2\log n - \tfrac{1}{2}l_2 n + l_3 n + a + c\right]\right),$$

where Y is a chi-squared random variable with 2 degrees of freedom. Thus

$$\lim EC_n = \lim(\log n)^{1/2}/l_2 n = \infty.$$

On the other hand, clumping turns out not to be a factor in the asymptotic distribution. Points large enough to lead to clumps containing more than one exceedance of $2r_1$ are rare enough to be irrelevant asymptotically. This lack of clumping seems to be a consequence of the rapid decay of normal distribution tails. With a heavier tailed distribution we expect clumping to be a factor in the asymptotics.

In Section 2 we eliminate some possible sources of exceedances. Consider four possible radii: r_1 given by (1.2),

$$(1.4)\quad r_2 = r_2(n) = \left[2\log n + \tfrac{1}{2}(k-3)l_2 n + 2(l_2 n + a(k) + c)\right]^{1/2},$$

$$(1.5)\quad r_3 = r_3(n) = \left[2\log n + kl_2 n\right]^{1/2}$$

and

$$r_0 = 2r_1 - r_2.$$

Although they appear different at first glance, all these radii and the points leading to exceedances are in a narrow annulus at $(2\log n)^{1/2} + O(l_2 n(\log n)^{-1/2})$. Possible exceedances are eliminated as follows. First, it is easy to see that there are no points X_i with $|X_i| > r_3$ with probability tending to 1. Then we show that with probability tending to 1, there are no pairs of points whose norms are both between r_2 and r_3 with the distance between them exceeding $2r_1$. Intuitively, points in this range are sparse enough so that their angular separations are not likely to be close enough to π to lead to a sufficiently large interpoint distance. Finally, we show that large interpoint distances with both points having norms less than r_2, and hence between r_0 and r_2, are also nonexistent with probability tending to 1.

With these cases eliminated, in Section 3 we study large interpoint distances based on one point inside the sphere of radius r_2 and another whose norm is between r_2 and r_3. Here we use a Poisson limit theorem for U-statistics from Silverman and Brown (1978). The proof of Theorem 1 is collected from those parts at the end of Section 3. Section 4 contains some final remarks.

For simplicity, calculations are done with O and o terms. At the cost of extra space, these can be replaced by bounds from below and above.

2. Elimination of some cases. We are concerned with pairs of points satisfying $|X_i - X_j| \geq 2r_1$. In this section we eliminate from consideration three cases. The following statements are shown to be true with probability approaching 1 as $n \to \infty$. First, there are no points with norms exceeding r_3. Second, pairs of points both having norms between r_2 and r_3 do not lead to any exceedances of $2r_1$. Third, pairs of points both within the sphere of radius r_2 do not lead to any exceedances of $2r_1$.

By introducing polar coordinates with r^2 denoting norm squared and with θ denoting angle, one can rewrite the standard normal density as

$$2^{-k/2} r^{k-2} e^{-r^2/2}\, dr^2\, d\theta / \Gamma(k/2),$$

where $d\theta$ corresponds to the uniform probability distribution on the unit sphere in R^k.

Consider first the number of points with $|X_i|^2 \geq r_3^2$. This has mean

$$\frac{n}{2^{k/2}\Gamma(k/2)} \int_{r_3^2}^{\infty} x^{k/2-1} e^{-x/2}\, dx. \tag{2.1}$$

For repeated use we cite a standard formula [see formula 8.357 of Gradshteyn and Ryzhik (1980)] for the incomplete gamma function:

$$\int_x^{\infty} e^{-t} t^{\alpha-1}\, dt = x^{\alpha-1} e^{-x}\big(1 + O(x^{-1})\big) \tag{2.2}$$

as $x \to \infty$.

Proposition 1. *The number of interpoint distances exceeding $2r_1$ involving at least one point whose norm is bigger than r_3 converges to 0 in probability as $n \to \infty$.*

Proof. From (2.2) we see that (2.1) is $O((\log n)^{-1})$. Thus the number of points with norm larger than r_3, and thus the number of exceedances of $2r_1$ involving them, converges to 0 in probability. □

Before covering the two cases remaining, we give some preliminary material used in both. We consider the distribution of the angle between independent points, the conditional distribution of $|X|^2$ given that $r_0 \le |X| \le r_2$ or $r_2 \le |X| \le r_3$ and the relationship between $|X_i - X_j|$ and the angle between X_i and X_j.

PROPOSITION 2. *Suppose that U and V are independent random variables uniformly distributed on the unit sphere in R^k. If*

$$\psi = \cos^{-1}(U \cdot V)$$

is the angle between U and V, then the distribution of ψ is symmetric about $\pi/2$ and $\cos^2 \psi$ has the beta distribution $B(1/2, (k-1)/2)$.

This fact is well known. See, for example, Muirhead [(1982), Section 1.5] for the relevant distributional relationships.

PROPOSITION 3. *Let X and Y be k-dimensional normal random vectors with zero means and identity covariance matrices. Let $w = |X|^2 - r_2^2$ and $z = r_2^2 - |Y|^2$. The conditional densities of w and z are, as $n \to \infty$,*

$$f_1\big(w | r_2^2 \le |X|^2 \le r_3^2\big) = \tfrac{1}{2} e^{-w/2}[1 + o(1)], \qquad 0 < w < r_3^2 - r_2^2, \tag{2.3}$$

$$f_2\big(z | r_0^2 \le |Y|^2 \le r_2^2\big) = \tfrac{1}{2} e^{-(r_2^2 - r_0^2 - z)/2}[1 + o(1)], \qquad 0 < z < r_2^2 - r_0^2. \tag{2.4}$$

PROOF. The conditional density f_1 for $0 < w < r_3^2 - r_2^2$ has the form

$$\frac{\big(w + r_2^2\big)^{k/2-1} e^{-(w + r_2^2)/2}}{\int_{r_2^2}^{r_3^2} \big(t + r_2^2\big)^{k/2-1} e^{-(t + r_2^2)/2}\, dt}. \tag{2.5}$$

The desired asymptotic formula follows from (2.2) applied to the integral in the denominator. The derivation of (2.4) is similar. □

PROPOSITION 4. *Suppose U and V are independently uniformly distributed on the spheres of radii $(r_2^2 + x)^{1/2}$ and $(r_2^2 + y)^{1/2}$. If x and y both are of order $O(l_2 n)$ and*

$$\big(r_2^2 + x\big)^{1/2} + \big(r_2^2 + y\big)^{1/2} \ge 2r_1,$$

then

$$\begin{aligned} P(|U - V| \ge 2r_1) &= \frac{1}{(k-1)\beta(1/2, (k-1)/2)} \\ &\quad \times \left(\frac{2(l_3 n + a + c) + x + y + O\big((x^2 + y^2)/\log n\big)}{\log n}\right)^{(k-1)/2} \\ &\quad \times \left(1 + O\left(\frac{l_2 n}{\log n}\right)\right). \end{aligned} \tag{2.6}$$

PROOF. Let ψ denote the angle between U and $-V$. The inequality $|U - V| \geq 2r_1$ is equivalent to

$$\cos\psi \geq 1 - \frac{(|U| + |V|)^2 - 4r_1^2}{2|U||V|} = 1 - \frac{t_1}{t_2}.$$

Because

$$|U| = r_2 + \frac{x}{2r_2} + O\left(\frac{x^2}{r_2^3}\right)$$

and

$$|V| = r_2 + \frac{y}{2r_2} + O\left(\frac{y^2}{r_2^3}\right),$$

one has

$$t_2 = 2r_2^2 + x + y + O\left(\frac{l_2^2 n}{\log n}\right) = 4\log n\left(1 + O\left(\frac{l_2 n}{\log n}\right)\right)$$

and

$$t_1 = 4(l_3 n + a + c) + 2(x + y) + O\left(\frac{x^2 + y^2}{\log n}\right).$$

It follows from Proposition 2 that

$$P\left(\cos^2\psi \geq \left(1 - \frac{t_1}{t_2}\right)^2, \psi < \pi/2\right)$$
$$= \frac{1}{2\beta(1/2,(k-1)/2)}\int_{(1-t_1/t_2)^2}^{1} z^{-1/2}(1-z)^{(k-3)/2}\,dz.$$

The formula (2.6) follows by standard asymptotic analysis using the formulas for t_1 and t_2. □

Now we consider interpoint distances exceeding $2r_1$ arising from pairs of points whose norms are both between r_2 and r_3.

PROPOSITION 5.

$$P\big(\textit{there exist } i, j,\ 1 \leq i < j \leq n, \textit{ with } r_2 \leq |X_i|, |X_j| \leq r_3 \cap |X_i - X_j| \geq 2r_1\big)$$
$$= O\left(\frac{(l_3 n)^{(k-1)/2}}{(l_2 n)^2}\right).$$

PROOF. Let p denote the probability that $r_2 \le |X_1| \le r_3$. Then

$$p = \frac{1}{2^{k/2}\Gamma(k/2)} \int_{r_2^2}^{r_3^2} x^{k/2-1} e^{-x/2}\, dx,$$

and the expected number of pairs X_i, X_j, $1 \le i < j \le n$, satisfying the individual norm requirements of the proposition is $\binom{n}{2} p^2$.

Conditional on $|X_i|$ and $|X_j|$, the likelihood that the points in such a pair are at least $2r_1$ apart is given by Proposition 4. If we integrate over the exact conditional density (2.5) of X_1 and X_2 and multiply by the expected number $\binom{n}{2} p^2$ of such pairs, we obtain the expected number of pairs exceedng $2r_1$ as

$$\frac{n(n-1)}{2^{k+1}\Gamma^2(k/2)(k-1)\beta(1/2,(k-1)/2)}$$
$$\times \int_{x=0}^{r_3^2-r_2^2} \int_{y=0}^{r_3^2-r_2^2} \left(r_2^2 + x\right)^{k/2-1} \left(r_2^2 + y\right)^{k/2-1} e^{-1/2[r_2^2+x+r_2^2+y]}$$
$$\times \left(\frac{2(l_3 n + a + c) + x + y + O\big((x^2+y^2)/\log n\big)}{\log n} \right)^{(k-1)/2}$$
$$\times \left(1 + O\left(\frac{l_2^2 n}{\log n}\right)\right) dy\, dx.$$

It is straightforward to check that this expression has the indicated asymptotic magnitude as $n \to \infty$. □

Finally consider interpoint distances exceeding $2r_1$ from pairs of points whose norms are both smaller than r_2 and hence are between r_0 and r_2.

PROPOSITION 6.

$$P\big(\textit{there exist } i \textit{ and } j,\ 1 \le i < j \le n,$$
$$\textit{with } \{|X_i| \le r_2\} \cap \{|X_j| \le r_2\} \cap \{|X_i - X_j| \ge 2r_1\}\big) = O\left(\frac{l_3 n}{l_2 n}\right).$$

PROOF. We proceed as with Proposition 5 except that now the restriction $|X_i - X_j| \ge 2r_1$ puts a constraint on $|X_i|$ and $|X_j|$. If $|X_i| = \sqrt{r_2^2 - x}$ and $|X_j| = \sqrt{r_2^2 - y}$, then $|X_i| + |X_j| \ge 2r_1$ implies

$$2r_2 - \frac{(x+y)}{2r_2} + O\left(\frac{x^2+y^2}{r_2^3}\right) \ge 2r_1$$

or

$$y \le 2(l_3 n + a + c) + O\left(\frac{l_2^2 n}{(\log n)^{1/2}}\right) - x.$$

Let s denote this upper limit on y. As in Proposition 5, the integral for the expected number of pairs $2r_1$ or more apart is

$$\frac{n^2}{2^{k+1}\Gamma^2(k/2)(k-1)\beta(1/2,(k-1)/2)}$$
$$\times \int_{x=0}^{r_2^2-r_0^2}\int_{y=0}^{s}\left(r_2^2-x\right)^{k/2-1}\left(r_2^2-y\right)^{k/2-1} e^{-1/2[r_2^2-x+r_2^2-y]}$$
$$\times \left(\frac{2(l_3 n + a + c) - x - y + O\left(l_2^2 n/\log n\right)}{\log n}\right)^{(k-1)/2} (1+o(1))\, dy\, dx.$$

Let $u = 2(l_3 n + a + c) - x - y$ and $v = x - y$. The preceding integral becomes

$$\frac{K}{(l_2 n)^2}\int_{u=o(1)}^{2(l_3 n+a+c)+o(1)}\int_{v=-[2(l_3 n+a+c)-u]}^{2(l_3 n+a+c)-u} \exp\{l_3 n + a + c - u/2\}$$
$$\times (u + o(1))^{(k-1)/2}\, dv\, du$$
$$= O\left(\frac{l_3 n}{l_2 n}\right). \qquad \square$$

3. The interesting case. We consider interpoint distances exceeding $2r_1$ where one point has a norm between r_2 and r_3 and the norm of the other is less than r_2. We will use a Poisson limit theorem for U-statistics [Silverman and Brown (1978)] to show that the number of such interpoint distances has asymptotically a Poisson distribution. This, combined with the results of the previous section, is used to give a proof of Theorem 1 at the end of this section.

The main result of this section is the following.

PROPOSITION 7. *Let*

$$c_n = \Big\{(i,j)\colon 1 \le i,j \le n, \big\{|X_i| \le r_2 \le |X_j| \le r_3 \text{ or } |X_j| \le r_2 \le |X_i| \le r_3\big\}, \text{ and } |X_i - X_j| \ge 2r_1\Big\}.$$

Then $|c_n|$ converges in distribution to a Poisson random variable with parameter e^{-c}.

PROOF. We use Theorem A of Silverman and Brown (1978), which is a Poisson limit theorem for U-statistics. Note that $|c_n|$, effectively a sum of symmetric indicator variables over pairs (i, j), is a U-statistic. Thus by Theorem A it suffices to show that

$$\binom{n}{2} P((1,2) \in c_n) \to e^{-c} \tag{3.1}$$

and

$$\binom{n}{3} P(\{(1,2) \in c_n\} \cap \{(1,3) \in c_n\}) \to 0 \text{ as } n \to \infty. \tag{3.2}$$

First consider (3.1):

$$\begin{aligned} &\binom{n}{2} P((1,2) \in c_n) \\ &\quad = 2\binom{n}{2} P(r_2 \le |X_1| \le r_3) P(\{|X_1 - X_2| \ge 2r_1\} \\ &\qquad \cap \{|X_2| \le r_2\} | r_2 \le |X_1| \le r_3). \end{aligned} \tag{3.3}$$

The first probability is

$$\begin{aligned} &\int_0^{r_3^2 - r_2^2} \frac{(r_2^2 + x)^{k/2-1} \exp\{-(r_2^2 + x)/2\}}{2^{k/2}\Gamma(k/2)} \, dx \\ &\quad = \frac{(\log n)^{k/2-1}}{\Gamma(k/2)} \exp\left\{-\frac{1}{2}\left(2 \log n + \left(\frac{k-3}{2}\right) l_2 n + 2(l_3 n + a + c)\right)\right\} \\ &\qquad \times \int_{x=0}^{K l_2 n} e^{-x/2} \frac{dx}{2} \\ &\quad = \frac{e^{-a-c} (\log n)^{(k-1)/4}}{n \Gamma(k/2) l_2 n} (1 + o(1)). \end{aligned} \tag{3.4}$$

The conditional probability in (3.3) can be computed as follows. Equation (2.3) gives the conditional density of X_1 given it is in the required range. Equation (2.6) of Proposition 4 gives the relevant conditional probability given both $|X_1|$ and $|X_2|$. Thus the conditional probability in (3.3) is the double integral of (2.6) against the conditional density of $|X_1|$ and the unconditional density of $|X_2|$. For later use in proving (3.2) we first calculate the integral with only the density of $|X_2|$. Call the result $\lambda(x)$.

It follows from Proposition 4 that

$$\begin{aligned} \lambda(x) = &\int_0^{2(l_3 n + a + c) + x + o(1)} (r_2^2 - y)^{k/2-1} e^{(r_2^2 - y)/2} \\ &\times \left[\frac{2(l_3 n + a + c) + x - y + o(1)}{\log n}\right]^{(k-1)/2} dy \\ &\times \left[2^{k/2} \Gamma\left(\frac{k}{2}\right) \beta\left(\frac{1}{2}, \frac{k-1}{2}\right)(k-1)\right]^{-1} (1 + o(1)). \end{aligned}$$

Setting $u = 2(l_3 n + a + c) + x - y$, we obtain

$$\lambda(x) = \frac{2^{(k-3)/2} e^{x/2}(1 + o(1))}{n\sqrt{\pi}(\log n)^{(k-1)/4}\Gamma((k+1)/2)}$$

(3.5) $$\times \int_{u=o(1)}^{x+2l_3 n + o(1)} e^{-u/2}(u + o(1))^{(k-1)/2}\, du$$

$$= \frac{2^{(k-3)/2} e^{x/2}}{n\sqrt{\pi}(\log n)^{(k-1)/4}}(1 + o(1)).$$

Now we integrate against the conditional density of $|X_1|$ to get

$$\frac{1}{2}\int_0^{r_3^2 - r_2^2} e^{-x/2}[\lambda(x)](1 + o(1))\, dx$$

$$= \frac{2^{(k-3)/2}}{n\sqrt{\pi}(\log n)^{(k-1)/4}}\left(\frac{k-1}{4}\right) l_2 n(1 + o(1)).$$

Multiplying these formulas, one obtains (3.1).

Computation for (3.2) is similar. The event $\{(1,2) \in c_n\} \cap \{(1,3) \in c_n\}$ can happen in two different ways, depending on whether $|X_1| \ge r_2$ or not. Consider first the case $r_2 \le |X_1| \le r_3$. Then

(3.6) $$P(\{(1,2) \in c_n\} \cap \{(1,3) \in c_n\} | r_2 \le |X_1| \le r_3) = \tfrac{1}{2}\int_0^{r_3^2 - r_2^2} e^{-x/2}\lambda^2(x)(1 + o(1))\, dx = O\big(n^{-2}(\log n)^{-(k-1)/4)}\big).$$

Multiplying by n^3 and $P(r_2 \le |X_1| \le r_3)$ gives a term of order $O((l_2(n))^{-1})$ for (3.2).

Next consider the case $|X_1| < r_2$. The probability that $(1,2) \in c_n$ and $(1,3) \in c_n$ is bounded above by the product

$$P(\{|X_1| \le r_2\} \cap \{|X_1 - X_2| \ge 2r_3\} | r_2 \le |X_2| \le r_3) \times P(r_2 \le |X_2| \le r_3)$$

$$\times P(r_2 \le |X_3| \le r_3) \times P\left(\cos\frac{\phi}{2} \ge 1 - \frac{(r_2 + r_3)^2 - 4r_1^2}{2r_2 r_3}\right),$$

where ϕ is the angle between X_2 and X_3. Straightforward geometry, that is, argument as in the proof of Proposition 4, shows that this condition is necessary for there to be any point of magnitude at most r_2 that is at least $2r_1$ from both X_2 and X_3.

The first term, computed previously, is $O(l_2 n / n(\log n)^{(k-1)/4})$. The next two are each $O((\log n)^{(k-1)/4}/n l_2 n)$. For the final term, the formula

$$\cos 2x = 2\cos^2 x - 1$$

implies

$$\cos \psi \ge 1 - (k+3)\frac{l_2 n}{\log n}(1 + o(1)),$$

or

$$\cos^2 \psi \geq 1 - 2(k+3)\frac{l_2 n}{\log n}(1+o(1)).$$

Using Proposition 2, we see that this event has probability at most

$$(3.7)\quad \begin{aligned} &\frac{1}{\beta(1/2,(k-1)/2)} \int_{1-2(k+3)l_2 n/\log n(1+o(1))}^{1} u^{-1/2}(1-u)^{(k-3)/2}\,du \\ &\quad = O\left(\left(\frac{l_2 n}{\log n}\right)^{(k-1)/2}\right) \end{aligned}$$

and the product of these probabilities has order $O((l_2 n)^{(k-3)/2}/n^3(\log n)^{(k-1)/4}$.

This completes the proof of Proposition 7. □

The proof of Theorem 1 is easy to collect now. Propositions 1, 5 and 6 imply that $C_n = |c_n| + o_p(1)$ as $n \to \infty$; Proposition 7 implies that C_n therefore is asymptotically Poisson distrbuted with parameter e^{-c}.

4. Concluding remarks. Instead of interpoint distances $|X_i - X_j|$ one could consider interpoint averages $|X_i + X_j|/2$. An examination of the proof of Theorem 1 shows that it also holds for card$\{(i,j)$: $1 \leq i < j \leq n$, $|X_i + X_j| < 2r_1\}$.

The observation that asymptotic clumping does not occur can be made rigorous. The proof of Theorem 1 shows that for any fixed j with probability tending to 1 as $n \to \infty$, the j largest interpoint distances involve $2j$ distinct points. This observation gives an informal test for outliers in a spherically symmetric normal point cloud. Any point that is involved in several of the largest interpoint distances could be flagged as a potential outlier.

Returning to the gun quality control problem that motivated this work, a practical question is whether the distribution of the range is well approximated by an extreme value distribution with appropriate moments. We have performed simulations that suggest two things. First, the distribution of the range is not well approximated by an extreme value distribution for small sample sizes. Second, the distributions of the squared range is very well approximated by the extreme value distribution, even for small samples. The fact that extreme value distribution approximations are often more accurate for squares of extreme order statistics is actually well known in the classical asymptotic theory of normal order statistics [see Haldane and Jayakar (1963)]. The case discussed in the introduction has $n = 10$ and $k = 2$. Figure 1 is a $Q - Q$ plot of the sample quantiles of $M^2(10)$ in two dimensions versus the quantiles of the standard extreme value distribution. This plot is based on 1000 replications. In the 1000 simulated squared ranges, $M^2(10)$ had average 14.75 and standard deviation 5.97. Using these moments (or the moments from a larger future simulation) and the extreme value distribution, it is

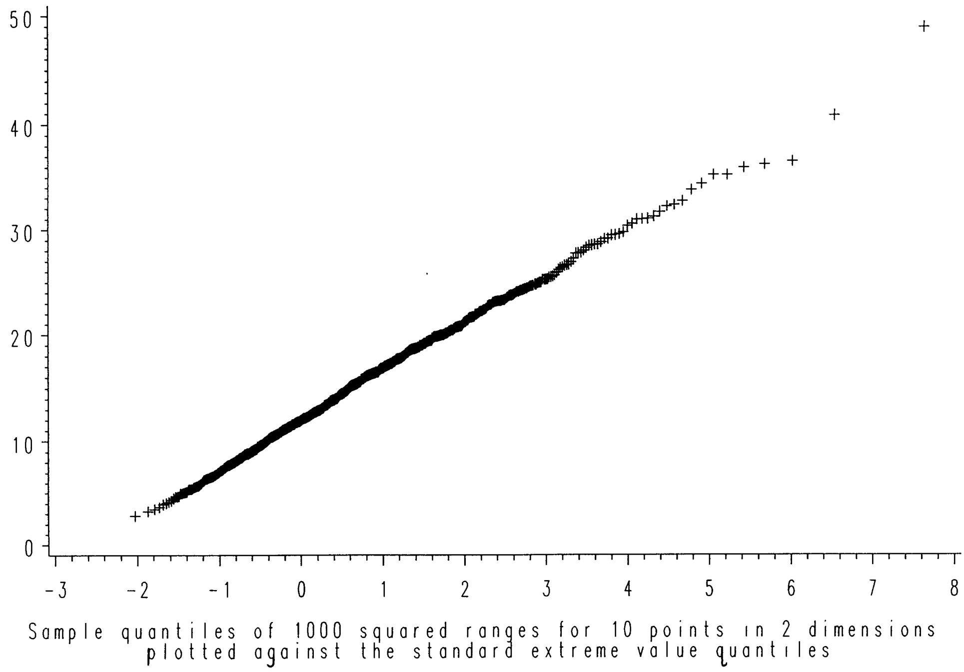

FIG. 1.

straightforward to do statistical inference about an unknown variance σ^2 based on $M^2(10)$.

Note that the asymptotics do a poor job of predicting moments. The asymptotic result of Corollary 1 suggests a mean of 9.15 and a standard deviation of 5.13 for $M^2(10)$ in two dimensions. This type of difference could be expected based on the way the results were derived. Asymptotically, pairs of points of all the types considered in Section 2 will not lead to the maximal interpoint distance. However, these types of pairs can have an effect for any reasonable n. Ignoring these possibilities is undoubtably a major factor in the difference in means. We have done simulations for samples as large as $n = 300$. The asymptotic moment approximations do not fit much better there. This is to be expected, as the error terms in Section 2 go to zero quite slowly, as slowly as $O(l_2^{-1} n l_3 n)$.

Acknowledgments. We would like to thank the referees and Associate Editor for helpful and thorough reviews of this paper. They pointed out the article by Silverman and Brown (1978), which simplified the proof of Theorem 1 significantly.

REFERENCES

ALDOUS, D. (1989). *Probability Approximation via the Poisson Clumping Heuristic*. Springer, New York.

BARBOUR, A. D. and EAGLESON, G. K. (1984). Poisson convergence for dissociated statistics. *J. Roy. Statist. Soc. Ser. B* **46** 397–402.

BARNETT, V. and LEWIS, T. (1984). *Outliers in Statistical Data*, 2nd ed. Wiley, New York.

CACOULLOS, T. and DECICCIO, H. (1967). On the distribution of the bivariate range. *Technometrics* **9** 476–480.

DANIELS, H. (1952). The covering circle of a sample from a circular normal distribution. *Biometrika* **39** 137–143.

DAVID, H. A. (1981). *Order Statistics*, 2nd. ed. Wiley, New York.

GRADSHTEYN, I. S. and RYZHIK, I. M. (1980). *Table of Integrals, Series, and Products*. Academic, New York.

GRUBBS, F. E. (1964). *Statistical Measures of Accuracy for Riflemen and Missile Engineers*. Edwards Broths, Ann Arbor, MI.

HALDANE, J. B. and JAYAKAR, S. D. (1963). The distribution of extremal and nearly extremal values in samples from a normal distribution. *Biometrika* **50** 89–94.

HENZE, N. (1987). On the fraction of random points with specified nearest-neighbor interrelations and degree of attraction. *Adv. in Appl. Probab.* **19** 873–895.

JAMMALAMADAKA, S. R. and JANSON, S. (1986). Limit theorems for a triangular scheme of U-statistics with applications to inter-point distances. *Ann. Probab.* **14** 1347–1358.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

SILVERMAN, B. and BROWN, T. (1978). Short distances, flat triangles and Poisson limits. *J. Appl. Probab.* **15** 815–825.

SIOTANI, M. (1959). The extreme value of generalized distance of the individual points in the multivariate normal sample. *Ann. Inst. Statist. Math.* **10** 183–208.

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MARYLAND
BALTIMORE COUNTY CAMPUS
BALTIMORE, MARYLAND 21228