Z. Wahrscheinlichkeitstheorie verw. Gebiete
68, 415-424 (1985)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete


# Evaluating Inclusion Functionals for Random Convex Hulls

Nicholas P. Jewell[1] and Joseph P. Romano[2]

[1] Program in Biostatistics, University of California, Berkeley, CA 94720, USA
[2] Department of Statistics, University of California, Berkeley, CA 94720, USA

**Summary.** The inclusion functional of a random convex set, evaluated at a fixed convex set $K$, measures the probability that the random convex set contains $K$. This functional is an analogue of the complement of the distribution function of an ordinary random variable. A methodology is described for evaluating the inclusion functional for the case where the random convex set is generated as the convex hull of $n$ i.i.d. points from a distribution function $F$ in the plane. For general $K$ and $F$, the inclusion probability is difficult to compute in closed form. The case where $K$ is a straight line segment is examined in detail and, in this situation, a simple answer is given for an interesting class of distributions $F$.

## 1. Introduction

Recently there has been great interest in the formulation of a general theory of random sets. Kendall [9] and Matheron [10] both provide a definition of a random set via a measure on a space of sets. Let $\mathscr{F}$ denote the collection of closed subsets of $\mathbb{R}^k$ and $\Sigma$ denote the usual Borel $\sigma$-field of subsets of $\mathscr{F}$. A random set $X$ is defined as a measurable map from an abstract probability space $(\Omega, \mathscr{B}, \mathscr{P}')$ into $(\mathscr{F}, \Sigma)$ with $\mathscr{P}$ being the induced probability measure on $\Sigma$.

For such random sets Eddy and Trader [4] introduced the inclusion function $G_X$ given by

$$G_X(K) = \Pr(K \subseteq X) \quad \text{for } K \in \mathscr{F}.$$

For random sets this function plays the role of the complement of the distribution function of an ordinary random variable. In particular $0 \leqq G_X \leqq 1$, $G(\emptyset) = 1$, $G$ is decreasing ($K_1 \subseteq K_2 \Rightarrow G(K_1) \geqq G(K_2)$) and $G$ is lower semi-continuous. In addition, under certain consistency requirements, the inclusion functional uniquely determines the probability measure $\mathscr{P}$. In fact, knowledge of the functional on the set of compact subsets of $\mathbb{R}^d$ is enough to determine $\mathscr{P}$ uniquely. This follows from Choquet's theorem (Theorem 2.2.1 in [10]) and the fact that

$G_X(K)=1-T_{X^c}(K)$ where $T_X(K)=\Pr(X\cap K\neq\emptyset)$ and $X^c$ is the complement of $X$. See also Theorem 5.1 and Corollary 5.3 of [4].

In this paper we provide a recipe for evaluating $G_X(K)$, at least in integral form, for the important case where the random set $X$ is generated as the convex hull of $n$ points, $p_1,\dots,p_n$, in the plane, independently and identically distributed according to a bivariate distribution $F$. That is, we show how to give expressions for the probability that the convex hull of $p_1,\dots,p_n$ contains any fixed compact set $K$. For general $K$ and $F$ the expression for this probability is not easy to compute in closed form, but we give a simple answer for the special case of $K$ a straight line segment and an interesting class of distributions $F$.

Finally, we indicate the importance of this problem, for $K$ a polygon, in computing moments of the (random) number of extreme points of the convex hull of $p_1,\dots,p_n$. Determining such quantities is crucial in developing the statistical properties of data analytic techniques based on convex hulls. For references to the uses of convex hulls in statistics see [6].

## 2. The Probability that the Convex Hull of $p_1,\dots,p_n$ Contains $K$

In this section we show that the problem of evaluating the probability that the convex hull of $p_1,\dots,p_n$ contains a fixed compact set is equivalent to a coverage problem in geometrical probability. Write $\mathrm{co}(p_1,\dots,p_n)$ for the convex hull of $p_1,\dots,p_n$.

Jewell and Romano [7] considered the following two problems:

i) drop $n$ points $p_1=(x_1,y_1),\dots,p_n=(x_n,y_n)$ in the plane independently and at random according to the bivariate distribution $F(x,y)$. Find the probability $G_F^n$ that a fixed disc is contained in $\mathrm{co}(p_1,\dots,p_n)$;

ii) let $H(\ell,m)$ be a bivariate distribution on $[0,\pi]\times[0,2\pi]$. Place $n$ random arcs on the circle of circumference $2\pi$ where the lengths $\ell_1,\dots,\ell_n$ and midpoints $m_1,\dots,m_n$ of the $n$ arcs are chosen according to a random sample of $n$ independent observations $(\ell_1,m_1),\dots,(\ell_n,m_n)$ drawn from the distribution $H$. Find the probability $S_H^n$ that the circumference of the circle is completely covered by the $n$ arcs.

The two problems are apparently unrelated but, in [7], it was shown that they are equivalent in the sense that if $H$ is prescribed, then $S_H^n=G_{H^*}^n$ where $H^*$ is a bivariate distribution in the plane derived from $H$ via a simple transformation. Conversely if $F$ is prescribed then $G_F^n=S_{F^*}^n$ where $F^*$ is a bivariate distribution on $[0,\pi]\times[0,2\pi]$ derived from $F$. Subsequently, in [7], a general integral formula was given for $G_F^n$ for any distribution $F$ thus solving both problems.

Now let $K$ be any fixed compact convex set in the plane and consider the more general case of problem (i) where we wish to determine the probability $G_F^n(K)$ that $K\subseteq\mathrm{co}(p_1,\dots,p_n)$. Thus $G_F^n$ above is $G_F^n(K)$, for $K$ a disc in the plane. We now show that the problem of evaluating $G_F^n(K)$ is also equivalent to problem (ii).

First, consider the support function of a convex set in the plane. The support function, $b_K$, of a compact convex set $K$ in the plane is a continuous function on the unit circle defined by

$$b_K(\gamma)=\sup_{\mathbf{k}\in K}\{\mathbf{k}\cdot\mathbf{e}_\gamma:\mathbf{e}_\gamma \text{ is the unit vector in the direction } \gamma\}.$$

There is a one-to-one correspondence between the collection of compact convex sets in the plane and the collection of support functions. If $K$ is a single point $\mathbf{k}$ with polar coordinates $(r,\theta)$ then $b_K(\gamma)=r\cos(\gamma-\theta)$ for $\gamma\in[0,2\pi]$. Moreover, if $K$ is the convex hull of the set of points $\{\mathbf{k}_i\}$, then $b_K(\gamma)=\sup_i b_{\mathbf{k}_i}(\gamma)$; i.e., the support function of the convex hull is just the pointwise maximum of the support functions of the $\mathbf{k}_i$'s. For further details on support functions and their use in studying random convex sets we refer to [3] and the references given there.

A trivial, but important, consequence of the definition of a support function is that a convex set $K_1$ is contained in another convex set $K_2$ if and only if $b_{K_1}(\gamma)\leqq b_{K_2}(\gamma)$ for all $\gamma\in[0,2\pi]$. In particular, if $K_1$ is a fixed comapct convex set and $K_2$ is the convex hull of a random sample of $n$ points $(r_i,\theta_i)$, $i=1,2,\ldots,n$, then the condition that $K_2$ contains $K_1$ is

$$b_{K_1}(\gamma)\leqq\max\{r_i\cos(\gamma-\theta_i): i=1,\ldots,n\} \quad \text{for all } \gamma\varepsilon[0,2\pi].$$

We also require some terminology borrowed from Rogers [12]. Let $K$ be a fixed compact convex set and suppose $P$ is some point exterior to $K$. Consider a line through $P$ which does not hit $K$. This line can be rotated in a clockwise sense about $P$ until it hits $K$. This (unique) line is called the clockwise critical line (CCL) from $P$ to $K$. If the CCL is oriented from $P$ to $K$ then $K$ will lie to the right of the CCL (see Fig. 1). Similarly there is a (unique) anticlockwise critical line (ACCL) oriented from $P$ to $K$ such that $K$ lies to the left of the ACCL.

Now draw both the CCL and ACCL from $P$ to $K$ and, without loss of generality, suppose the origin 0 is interior to $K$. To both the CCL and ACCL there correspond angles $\psi_1$ and $\psi_2$, respectively, formed by the perpendiculars $OH_1$ and $OH_2$ with the fixed direction $0x$ (see Fig. 1).

We will call $\psi_1$ the clockwise critical angle (CCA) and $\psi_2$ the anticlockwise critical angle (ACCA).

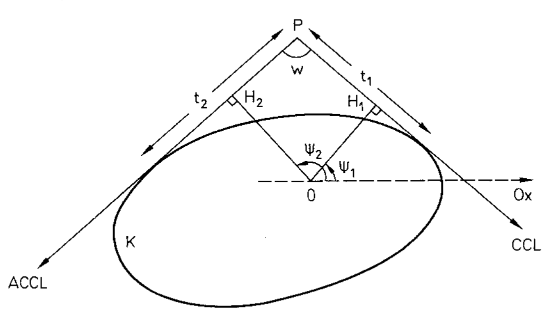


**Fig. 1**

**Lemma.** *Let $K$ be a compact convex set containing the origin and let $P=(r,\theta)$ be a point exterior to $K$. Then $\{\gamma: b_K(\gamma) \leqq r\cos(\gamma-\theta)\}$ is an interval, modulo $2\pi$. If the interval is considered as an arc on the unit circle then the arc's clockwise and anticlockwise endpoints are the CCA and ACCA respectively.*

*Proof.* Consider the set of all lines passing through $P$. These lines can be indexed by the angle $\psi$ formed by the perpendicular from the origin to the line with the fixed direction $0x$. Call the line indexed this way $L_\psi$. Thus the CCL is $L_{\psi_1}$ etc. It is clear that $b_K(\psi_1)=r\cos(\psi_1-\theta)$ and $b_K(\psi_2)=r\cos(\psi_2-\theta)$. If we mark the points $\psi_1$, $\psi_2$ on the unit circle, then for $\psi$ lying between $\psi_1$ and $\psi_2$(i.e, $\psi$ lies on the clockwise side of $\psi_2$ and the anticlockwise side of $\psi_1$) the line $L_\psi$ does not hit $K$. This follows from the definitions of the CCL and ACCL. In this situation $b_K(\psi)<r\cos(\psi-\theta)$. Alternatively for $\psi$ lying elsewhere on the unit circle the line $L_\psi$ passes through $K$ and then $b_K(\psi) \geqq r\cos(\psi-\theta)$. The convexity of $K$ requires that the set of $\psi$ with $b_K(\psi) \geqq r\cos(\psi-\theta)$ is a connected set, i.e., an interval (modulo $2\pi$).

Let $p_1, \ldots, p_n$ be chosen independently from the distribution $F$, and let $b_i(\gamma)=b_{p_i}(\gamma)$ for $i=1, \ldots, n$. For distributions $F$ with no support in $K$, the immediate consequence of the lemma and preceding comments is that the event $b_K(\gamma) \leqq \max_i \{b_i(\gamma)\colon i=1, \ldots, n\}$ for all $\gamma \in [0, 2\pi]$ is equivalent to the event that the circle of circumference $2\pi$ is covered by a random sample of $n$ arcs when the clockwise and anticlockwise endpoints of the arcs are identically and independently distributed according to a distribution function $H_0$. The distribution $H_0$ is the joint distribution of the CCA and ACCA relative to $K$ that is induced by a point $P$ chosen according to $F$. Note that $H_0$ depends on both $F$ and $K$.

In order to use the results of [7] we need the joint distribution $H(\ell, m)$ of the length and midpoints of the arcs rather than $H_0$. This is easily computed. Below we indicate how to evaluate $H$ from $F$ and $K$ and, in Sect. 4, the particular case when $K$ is a straight line segment will be explored in detail.

For distributions $F$ with no support in $K$ we have thus established that $G_F^n(K)=S_H^n$, a coverage probability, where $H$ is derived from $F$ and $K$ according to the above description. The value of $S_H^n$ can then be computed, in principal, from the results of [7] described at the beginning of this section. For distributions $F$ with support in $K$ we must replace $F(\mathbf{x})$ in the above with $F_K(\mathbf{x})$, the conditional distribution of $F$ given that $\mathbf{x} \notin K$. In this case we have

$$G_F^n(K)=\sum_{j=0}^{n} G_{F_K}^j(K)\, p^j(1-p)^{n-j}\binom{n}{j}$$

where $p=\Pr_F(\mathbf{x} \notin K)$.

## 3. The Derivation of $H$ from $F$ and $K$

The description of $H$ in terms of $F$ and $K$ depends on a construction of Crofton [2]. In particular, when $K$ has a smooth boundary, we give the Jacobian of the transformation $P=(x,y) \to (\psi_1, \psi_2)$ where $\psi_1, \psi_2$ are the CCA and ACCA of $P$ relative to $K$, respectively. The point $P=(x,y)$ is chosen according to $F$, and is assumed to be outside of $K$ (see Fig. 1). The angles $\psi_1, \psi_2$ uniquely determine the point $P$. We wish to express the distribution of $\psi_1, \psi_2$ in terms of $F$ and $K$.

Let $t_1, t_2$ be the distances from $P$ to the points of tangency of the CCL and ACCL, respectively. Let $w$ be the smaller angle formed by the CCL with the ACCL at $P$. Then it can be shown that $dx\,dy = (t_1 t_2/\sin w)\,d\psi_1\,d\psi_2$ (see [13, p. 26–27] for details).

There is no simple expression for the length $\ell$ and midpoint $m$ of the arc in terms of $\psi_1, \psi_2$ that applies for all possible values. However, it is easy to verify that $d\psi_1\,d\psi_2 = d\ell\,dm$. Also note that $w = \pi - \ell$, (see Fig. 1). Thus

$$dx\,dy = (t_1 t_2/\sin \ell)\,d\ell\,dm. \tag{1}$$

We note that (1) often holds for convex sets $K$ which do not possess tangents at every point of the boundary.

In addition to the Jacobian, we also need to express the coordinates $x, y$ (and $t_1, t_2$) in terms of $\ell, m$. This is complex and requires detailed knowledge of the boundary of $K$. For $K$ a disk or point, it is easy. The more complex case where $K$ is a straight line segment is described in Sect. 4.

To complete the evaluation of $H$ we must also describe the support of the distribution in $[0, \pi] \times [0, 2\pi]$. When the support of $F$ is $K^c$ and $K$ is a disk, the support is the whole rectangle. However, this need not be the case for more complex $K$. An example when the support of $H$ is never the whole rectangle is given when $K$ is a straight line segment (see Sect. 4).

There is an interesting class of distributions whose definition arises from this parametrization of the coordinates of $P$. Define the class of distributions $F$ with no support in $K$ to be *circularly symmetric about* $K$ if the marginal distribution of $m$ after the change in variables from $(x, y)$ to $(\ell, m)$ with Jacobian given by (1) is uniform on $[0, 2\pi]$. The mass of such distributions is uniformly spread "around" $K$ in terms of the CCA and ACCA. In the special case where $K$ is a single point this is just the class of distributions spherically symmetric about the point.

## 4. The Straight Line Segment Case

In this section we provide formulae for $G_F^n(K)$ when $K$ is a straight line segment and evaluate such expressions for a family of distributions $F$. In the conclusion we indicate the importance of the case where $K$ is a polygon in the determination of statistical properties of the convex hull of $p_1, \ldots, p_n$.

Initially, for simplicity, let $K$ be the straight line joining the points $P_1 = (1, 0)$ and $P_2 = (-1, 0)$. Let $F$ be a continuous distribution function on the plane. We wish to evaluate $G_F^n(K)$. Suppose $P$ is dropped in the plane according to $F$.

Although $K$ does not satisfy the conditions used in Sect. 3 it is easy to verify that (1) still holds. For example, if $y > 0$, the CCL is the line through $P$ and $P_1$. Then $1 - x = y \tan \psi_1$ and $1 + x = -y \tan \psi_2$. Differentiating both these equations with respect to $\psi_1$ and $\psi_2$, and solving for the appropriate partial derivatives yields (1). A similar analysis shows that (1) holds if $y < 0$.

We now investigate the boundaries in $\ell$-$m$ space which describe the image of the transformation of the coordinate system $(x, y)$ to the coordinate system $(\ell, m)$. The simplest approach is to describe the level sets of $\ell$ and $m$ in the plane. The level sets of $\ell$ correspond to the level sets of $\psi_2 - \psi_1$ and the level sets of $m$ correspond to level sets of $\psi_1 + \psi_2$.

Analysis of the level sets of $\psi_2 - \psi_1$ yields the following description of the level sets of $\ell$. For a fixed $\ell_0$, consider the two circles $ky^2 \pm 2y + kx^2 = k$ where $k = \tan \ell_0$. The locus of points $(x, y)$ which lead to constant $\ell = \ell_0$ is the part of the circle $ky^2 + 2y + kx^2 = k$ with $y > 0$ and the reflection of this part of the circle through the $O_x$ axis. The remaining arcs of the two circles correspond to the level set of $\ell = \pi - \ell_0$. This accounts for the level sets of $\ell$ for $0 < \ell < \pi$. Note that the level set of $\ell = \pi/2$ is just the circle with center the origin and radius 1. For $\ell = 0$ the level set is the line $y = 0$.

A similar analysis yields the level sets of $m$. Let $v = \tan 2m$ and consider the rectangular hyperbola, $vy^2 + 2xy - vx^2 + v = 0$, which passes through $P_1$ and $P_2$. A given value of $v$ corresponds to four different values of $m$. The four sections of the hyperbola according to whether $x \lesseqgtr 0$, $y \lesseqgtr 0$, represent the level sets corresponding to these four values of $m$.

The special case $v = 0$, $\pm\infty$ yields the level sets of $m$ for the values 0, $\pi/2$, $\pi$, $3\pi/2$, and $\pi/4$, $3\pi/4$, $5\pi/4$, $7\pi/4$. Points on the line joining $P_1$ and $P_2$ correspond to the value $m = \pi/2$ but $\ell = 0$.

It is straightforward to use the information on the level sets to show that for a fixed $\ell$, $m$ ranges over the two intervals $[(\pi - \ell)/2, (\pi + \ell)/2]$ and $[(3\pi - \ell)/2, (3\pi + \ell)/2]$. As $\ell$ moves from 0 to $\pi$ this describes the region, $\mathscr{R}$, of $\ell - m$ space which is the image of the transformation of the coordinate system $(x, y)$ to the coordinate system $(\ell, m)$. The region is illustrated in Fig. 2.

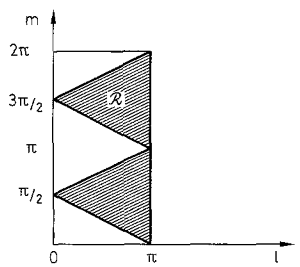


**Fig. 2.** The shaded region $\mathscr{R}$ corresponds to the image of the $x - y$ plane under the transformation $(x, y) \to (\ell, m)$

We return to the evaluation of $G_F^n(K)$. In Sect. 2, 3 we determined that $G_F^n(K) = S_H^n$ where $H$ is the joint distribution of $\ell, m$ induced by $F(x, y)$ in the prescribed manner. We also have

$$dH(\ell, m) = \frac{\sin \ell}{t_1 t_2} dF(x(\ell, m), y(\ell, m)). \tag{2}$$

and $H$ is supported on the region $\mathscr{R}$.

Now, in [7] it was shown that $S_H^n = G_{H^*}^n(D)$ where $D$ is the disk of unit radius center at the origin, and $H^*(R, \Theta)$ is the distribution on the plane with support outside $D$ derived from $H(\ell, m)$ via the following transformation

$$R(\ell, m)=\left[\sin\left(\frac{\pi-\ell}{2}\right)\right]^{-1}, \qquad \Theta(\ell, m)=m. \tag{3}$$

The support of $H^*$ is the image of $\mathscr{R}$ under this transformation, i.e., the set $\mathscr{R}^*$ of $(X, Y)$ in the plane such that $|Y| \geqq 1$.

From (2), (3) we have

$$dH^*(R, \Theta)=(R/2)(R^2-1)^{1/2}(\sin \ell/t_1 t_2)\, dF(x(R, \Theta), y(R, \Theta)).$$

With a little algebra, working from Fig. 1 with $K$ a line segment gives

$$t_1 t_2=R^2(R^2 \sin^2 \Theta-1)/(R^2-1)$$

on the support of $H^*$.

We now have a complete description of $H^*$ in terms of $F$ and we can provide a formula for $G^n_{H^*}(D)$ using the results of [7]. First we need some notation.

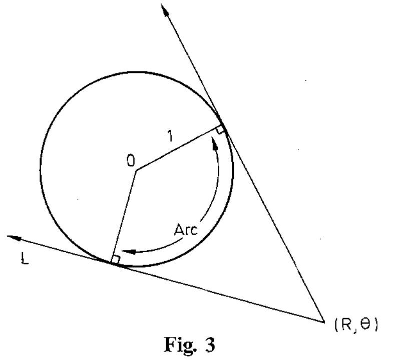


**Fig. 3**

Consider a point $(R, \Theta)$ in the plane, chosen according to $H^*$, and the corresponding arc on the unit circle supported by $(R, \Theta)$ as described in Fig. 3. For each point $(R_i, \Theta_i)$, let $L_i$ be the directed tangent from $(R_i, \Theta_i)$ to the unit disc with the disc on the right of $L_i$. Let $\mathscr{C}_k$ be the event that there are gaps after (in a clockwise sense) each of the first $k$ supported arcs, having selected only the first $k$ of $(R_i, \Theta_i)$ according to $H^*$. Let $C_k$ describe the $k$-dimensional region where $(R_1, \Theta_1), \ldots, (R_k, \Theta_k)$ must fall for $\mathscr{C}_k$ to occur. Given $\mathscr{C}_k$, let $\mathscr{B}_k$ be the event that the subsequent $(n-k)$ points do not cover any of those gaps, i.e., each of $(R_{k+1}, \Theta_{k+1}), \ldots, (R_n, \Theta_n) \in B_k = D_1 \cap \ldots \cap D_k$ where $D_i$ is the half-plane to the right of $L_i$. Then Theorem 4.1 of [7] gives

$$G^n_{H^*}(D)=1+\sum_{k=1}^{n}(-1)^k \binom{n}{k} \int_{C_k} \ldots \int \Big[\int_{B_k} dH^*\Big]^{n-k} dH^*(R_1, \Theta_1) \ldots dH^*(R_k, \Theta_k).$$

In general, when the support of $H^*$ is the complement of the disk, each event $\mathscr{C}_k$ occurs with positive probability. However, in the case under consideration, $H^*$ has support confined to the region $\mathscr{R}^*$ described above. With this constraint there can be gaps after at most the first two supported arcs, i.e., $\Pr_{H^*}(\mathscr{C}_k)=0$

for $k>2$. This is easily seen, since there cannot be a gap after each of two supported arcs, corresponding to points $(R,\Theta)$ with $R\sin\Theta\geqq 1$. The same holds if $R\sin\Theta\leqq -1$ for both points. Thus there can be gaps after at most two arcs, one corresponding to a point with $R\sin\Theta\geqq 1$, the other corresponding to a point with $R\sin\Theta\leqq -1$. Thus the above formula simplifies to:

$$\begin{aligned} G_F^n(K)=G_{H^*}^n(D)=1-n\int_{C_1}\Big[\int_{B_1} dH^*\Big]^{n-1} dH^*(R_1,\Theta_1) \\ +[n(n-1)/2]\int_{C_2}\int\Big[\int_{B_2} dH^*\Big]^{n-2} dH^*(R_1,\Theta_1)\,dH^*(R_2,\Theta_2). \end{aligned} \tag{4}$$

The parametrization of $B_1$, $B_2$, $C_1$ and $C_2$ is straightforward but complicated. Again there are simplifications due to the restriction of the support of $H^*$. Full details are given in a technical report from the University of California [8] which is available from the authors.

## Section 5. An Example

The formula (4) above, together with the appropriate parametrizations, give a formula for $G_F^n(K)$ when $K$ is the straight line segment joint $(-1,0)$ to $(1,0)$. Exact computation for finite and for general distribution $F$ is formidable. The main value of (4) may be in providing a method to evaluate asymptotic values of $G_F^n(K)$ as $n\to\infty$. Here, for a special class of distributions $F$, we evaluate $G_F^n(K)$ for finite $n$. This is the class of distributions which are circularly symmetric about a straight line segment according to the definition in Sect. 3. The densities are given by:

$$dF(x,y)=c(t_1^2\,t_2^2/|y|)\,f(\pi-w)\,dx\,dy$$

where $f$ is an arbitrary positive integrable function on $[0,\pi]$ and $c$ is the appropriate integrating constant. For a given $(x,y)$, the values of $t_1, t_2, w$ are as shown in Fig. 1 with $K$ the straight line segment joining $(-1,0)$ to $(1,0)$. In the coverage version of the evaluation of $G_F^n(K)$ we have $dH(\ell,m)=2cf(\ell)\,d\ell\,dm$ on $\mathscr{R}$, (see (2)). Thus the joint density only depends on $\ell$ and the marginal distribution of $m$ is uniform on $[0,2\pi]$. This makes the computation of $S_H^n=G_{H^*}^n(D)$ simpler. We compute $S_H^n$ for the simplest distribution in this class where $f$ is a constant function. Then

$$dF(x,y)=(1/2\pi^2)(t_1^2\,t_2^2/|y|)\,dx\,dy, \qquad (x,y)\in\mathbb{R}^2$$

and

$$dH(\ell,m)=1/\pi^2\,d\ell\,dm, \qquad (\ell,m)\in\mathscr{R}.$$

Thus the lengths and locations are uniformly distributed over the region $\mathscr{R}$. Note that the following computation of $S_H^n$ provides an example of a coverage problem where the lengths and locations of the arcs are not independent. This is the first such example for which explicit coverage probabilities have been evaluated for finite $n$. Examining the region $\mathscr{R}$ indicates that for this distribution $H$, the average arc length near location $m=0$ and $\pi$ is larger than near $m=\pi/2$ and $3\pi/2$.

We evaluate expression (4). Using the parametrization of $B_1$ and $B_2$ with a tedious amount of complex calculation yields:

$$\int_{B_1} dH^*(R,\Theta)=(3/4)-(\beta_1^2/\pi^2)$$

$$\int_{B_2} dH^*(R,\Theta)=((\beta_1+\beta_2)/\pi)-((\beta_1+\beta_2)^2/2\pi^2)$$

where $\beta_j=m_j-(\ell_j/2)$, $(j=1,2)$. For details see [8]. Then using the parametrization of $C_1$ and $C_2$ given in [8], we obtain

$$\int_{C_1}\Big[\int_{B_1} dH^*\Big]^{n-1} dH^*(R_1,\Theta_1)=\int_{C_1}((3/4)-(\beta_1^2/\pi^2))^{n-1}\, dH^*(R_1,\Theta_1)$$
$$=2^{-2(n-1)}\int_0^1(3-t^2)^{n-1}\,dt;$$

$$\int\int_{C_2}\Big[\int_{B_2} dH^*\Big]^{n-2} dH^*(R_1,\Theta_1)\,dH^*(R_2,\Theta_2)$$
$$=\int\int_{C_2}[((\beta_1+\beta_2)/\pi)-((\beta_1+\beta_2)^2/2\pi^2)]^{n-2}\,dH^*(R_1,\Theta_1)\,dH^*(R_2,\Theta_2)$$
$$=4\int_0^1 ds\int_0^2 t^2[(s+t)-((s+t)^2/2)]^{n-2}\,dt.$$

Hence

$$G_F^n(K)=S_H^n$$
$$=1-(n/2^{2(n-1)})\int_0^1(3-t^2)^{n-1}\,dt+2n(n-1)\int_0^1 ds\int_0^2 t^2[(s+t)-((s+t)^2/2)]^{n-2}\,dt.$$

Table 1 lists the value of this probability for small values of $n$ together with $S_{H_0}^n$, $S_{H_1}^n$ for comparison, where $dH_0(\ell,m)$ is the uniform distribution on $[0,\pi]\times[0,2\pi]$ and $dH_1(\ell,m)$ is the distribution of uniform locations $m$ and fixed arc lengths $\ell=2\pi/3$ which is the mean arc length under distribution $H$.

**Table 1.** Coverage probabilities

| $n$ | $S_H^n=G_F^n(K)$ | $S_{H_0}^n$ | $S_{H_1}^n$ |
|---|---|---|---|
| 3 | 0.017 | 0.030 | 0 |
| 4 | 0.069 | 0.092 | 0.037 |

## Section 6. Conclusion

We have shown how to compute $G_F^n(K)$ when $K$ is the straight line segment joining $P_1=(1,0)$ and $P_2=(-1,0)$. For general $P_1=(x_1,y_1)$, $P_2=(x_2,y_2)$ the above methods can be used following three simple transformations:

(i) translate so that $((x_1+x_2)/2, (y_1+y_2)/2$ becomes the origin;
(ii) rotate so that $\overrightarrow{P_2P_1}$ is parallel to the $0x$ axis;
(iii) shrink (or magnify) so that $|0P_1|=|0P_2|=1$.

The Jacobian of the composition of these transformations $\phi$ is $d^2/4$ where $d=|P_1P_2|$. Since the convex hull of $n$ points contains $P_2P_1$ if and only if the convex hull of the $n$ transformed points (after application of $\phi$) contains the line $T$ joining $(-1,0)$ to $(1,0)$, we have $G_F^n(\overrightarrow{P_2P_1})=G_{F^*}^n(T)$ where $dF^*$ is the distribution of $\phi(x,y)$.

Finally, we indicate an important application of the case where $K$ is a straight line segment. Let $N$ be the number of extreme points of the convex hull of $n$ points identically and independently distributed according to $F$. Let $q$ be the probability that the first point $p_1$ is not an extreme point of the convex hull. Then $E(N)=n(1-q)$. Formulae for $E(N)$ for fixed values of $n$ and in the limit, as $n\to\infty$, are given in [11, 5, 1]. At present no formulae are known for var($N$), even asymptotic. Note that $\mathrm{var}(N)=nq(1-nq)+n(n-1)q_2$ where $q_2$ is the probability that neither of the first two points $p_1, p_2$ are extreme, i.e., $p_1$ and $p_2\in\mathrm{co}(p_3,\ldots,p_n)$. In the current notation

$$q_2=\iint G_F^{n-2}(\overrightarrow{p_2p_1})\,dF(p_1)\,dF(p_2).$$

Although the difficulty of evaluating $G_F^{n-2}(\overrightarrow{p_2p_1})$ makes evaluation of $q_2$ for finite $n$ a formidable problem, it is hoped that the methods of evaluating $G_F^n(K)$ described in this paper will make an asymptotic analysis of $q_2$ and hence var($N$) possible. The same ideas can be extended to give formulae for the higher moments of the number of extreme points.

## References


1. Carnal, H.: Die konvexe Hülle von $n$ rotationssymmetrisch verteilten Punkten. Z. Wahrscheinlichkeitstheorie verw. Gebiete **15**, 168–179 (1970)
2. Crofton, M.W.: Probability. Encyclopaedia Brittanica, 9th ed., Vol. **19**, 768–788 (1885)
3. Eddy, W.F.: The distribution of the convex hull of a Gaussian sample. J. Appl. Probability **17**, 686–695 (1980)
4. Eddy, W.F., Trader, D.A.: Probability functionals for random sets. Technical Report No. 252, Department of Statistics, Carnegie-Mellon University, Pittsburgh (1982)
5. Efron, B.: The convex hull of a random set of points. Biometrika **52**, 331–342 (1965)
6. Green, P.J.: Peeling bivariate data, in *Interpreting Multivariate Data*, ed., V. Barnett, New York; Wiley, pp. 3–19, 1981
7. Jewell, N.P., Romano, J.P.: Coverage problems and random convex hulls. J. Appl. Probability **19**, 546–561 (1982)
8. Jewell, N.P., Romano, J.P.: The probability that a random convex hull contains a fixed set. Technical report, Program in Biostatistics, University of California, Berkeley (1982)
9. Kendall, D.G.: Foundation of a theory of random sets, in *Stochastic Geometry*, eds., E.F. Harding eds., E.F. Harding and D.G. Kendall, New York: Wiley, pp. 322–376 (1974)
10. Matheron, G.: *Random Sets and Integral Geometry*. New York: Wiley, 1975
11. Renyi, A., Sulanke, R.: Über die Konvexe Hülle von $n$ zufällig gewählten Punkten. Z. Wahrscheinlichkeitstheorie verw. Gebiete **2**, 75–84 (1963)
12. Rogers, L.C.G.: The probability that two samples in the plane will have disjoint convex hulls. J. Appl. Probability **15**, 790–802 (1978)
13. Solomon, H.: *Geometric Probability*. SIAM, Philadelphia (1978)


Received April 5, 1983