ISRAEL JOURNAL OF MATHEMATICS, Vol. 72, No. 3, 1990

THE EXPECTED VALUE OF SOME FUNCTIONS OF THE CONVEX HULL OF A RANDOM SET OF POINTS SAMPLED IN R^d

BY

ISAAC MEILIJSON

Raymond and Beverly Sackler Faculty of Exact Sciences, School of Mathematical Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel

ABSTRACT

This paper presents formulas and asymptotic expansions for the expected number of vertices and the expected volume of the convex hull of a sample of n points taken from the uniform distribution on a d-dimensional ball. It is shown that the expected number of vertices is asymptotically proportional to $n^{(d-1)/(d+1)}$, which generalizes Rényi and Sulanke's asymptotic rate $n^{(1/3)}$ for $d = 2$ and agrees with Raynaud's asymptotic rate $n^{(d-1)/(d+1)}$ for the expected number of facets, as it should be, by Bárány's result that the expected number of s-dimensional faces has order of magnitude independent of s. Our formulas agree with the ones Efron obtained for $d = 2$ and 3 under more general distributions. An application is given to the estimation of the probability content of an unknown convex subset of R^d.

Introduction

Let $X_1, X_2, \ldots, X_n$ (with $n \geq d + 1$) be independent and identically distributed d-dimensional random variables having a uniform distribution inside the d-dimensional unit ball S with center at the origin. Let A_n be the convex hull of the n points, H_n its volume, V_n the number of its vertices, T_n the number of its vertices that are at a local maximal distance from points in A_n to the center of S and let F_n be the number of facets ($(d-1)$-dimensional faces) of A_n.

We will present exact formulas for the expected values of H_n, V_n, and T_n in integral form, illustrate those of H_{d+1} and V_{d+2} in explicit form and we will show $E(V_n)$ and $E(T_n)$ to be asymptotically proportional (as $n \to \infty$) to $n^{(d-1)/(d+1)}$. The proportionality factors will be presented explicitly. The rate at which $E(H_n)$

Received December 19, 1989

approaches the volume $v_d = \pi^{(d/2)}/\Gamma((d/2)+1)$ of the unit ball is obtained from the relation (Efron [4])

$$E(V_n) = n[1 - E(H_{n-1})/v_d], \tag{1}$$

and the rate of growth of vertices satisfies

$$\lim_{d\to\infty}\lim_{n\to\infty}(2/d)n^{-(d-1)/(d+1)}E(V_n) = \lim_{d\to\infty}\lim_{n\to\infty}(2/d)n^{-(d-1)/(d+1)}E(T_n) = 1. \tag{2}$$

The formula for $E(H_{d+1})/v_d$, the expected proportion of the unit ball occupied by the random simplex, is well known to be

$$\frac{E(H_{d+1})}{v_d} = \frac{\pi\eta((d+1)^2)}{(\pi\eta(d+1))^{d+1}} \tag{3}$$

where

$$\eta(m) = \left(\int_0^{\pi/2}\sin^m x\,dx\right)^{-1}. \tag{4}$$

Rényi and Sulanke [11] obtained the asymptotic behavior of $E(V_n) = E(F_n)$ for points sampled from a uniform distribution on a compact convex set in R^2. If the set is a polygon, $E(F_n)$ and $E(V_n)$ are asymptotically proportional to $\log n$ while if the boundary of the set is smooth, the rate is $n^{(1/3)}$, as we obtain here. Efron [4] presents formulas for $E(V_n)$ and for $E(F_n) = 2E(V_n) - 4$ for $d = 3$. For $d \geq 4$ there is no clear relation between the number of facets and vertices of convex polytopes. Raynaud [10] claims, to the contrary, that $F_n = (d-1)V_n - (d+1)(d-2)$ with probability one, extending $F_n = V_n$ for $d = 2$ and $F_n = 2V_n - 4$ for $d = 3$. This claim is wrong and so is the result based on it. We present the correct asymptotic computation of $E(V_n)$.

Counterexamples to Raynaud's relation are provided by the neighborly polytopes (see Grünbaum [7]), i.e., those for which any $[(d/2)]$ vertices are contained in some facet; there is a strict relation

$$F = \binom{V - \frac{d}{2}}{\frac{d}{2}} + \binom{V - \frac{d}{2} - 1}{\frac{d}{2} - 1} \quad \text{or} \quad 2\binom{V - \left[\frac{d}{2}\right] - 1}{\left[\frac{d}{2}\right]} \tag{5}$$

(depending on whether d is even or odd) between the number of vertices V and the number of facets F of neighborly polytopes, which is linear only for $d = 2$ or 3. It is known [7] that there exist in every dimension neighborly polytopes of any number of vertices and that these maximize the number of facets among all polytopes with a given number of vertices.

We obtain formulas that might help in checking the extent to which the conjecture of Gale, Grünbaum, Motzkin and others, that "general" polytopes are neighborly or nearly so, holds for random convex polytopes. We restrict the asymptotic analysis to the evaluation of expectations. Groeneboom [6] has proved (for $d = 2$) that the number of vertices satisfies a central limit theorem.

The fact that the expected number (f_0, f_{d-1}) of zero-dimensional faces (vertices) and of $(d-1)$-dimensional faces (facets) are asymptotically proportional to each other is a strengthened special case of Bárány's result that, for any $0 < s \leq d-1$,

$$0 < \liminf_{n \to \infty} (f_s/f_0) \leq \limsup_{n \to \infty} (f_s/f_0) < \infty.$$

Schneider and Wieacker [12] relate the expected number of faces to the expected volume and to the rate at which the mean width of A_n approaches the diameter of the ball. For comprehensive surveys, consult Buchta [2] and Bárány [1].

This work was motivated by an estimation problem [8]. A sample from a known distribution in R^d is taken to estimate the location and probability content of an unknown convex set S in R^d. At each sample point we are told whether or not the point is in S. Obviously, the number (V_n) and location of the vertices of the convex hull A_n of the sample points inside S is of interest, but so is the number (C_n) and location of the "cusps" among the sample points outside S. A "cusp" is a sample point that does not belong to the "shade" of A_n beyond any other sample point outside S.

A variant of $E(C_n)$ will be computed, assuming a uniform distribution between two concentric balls, of radii 1 and $1 + \Delta$. The inner ball is S, and C_n counts the number of sample points that do not belong to the shade of S beyond some other sample point. $E(C_n)$ will be shown to be asymptotically proportional to $n^{(d-1)/(d+1)}$ as n tends to infinity. If the volume between the two concentric balls equals the volume of the smaller of the two balls, then (compare with (2))

$$\lim_{d \to \infty} \lim_{n \to \infty} (2/d) n^{-(d-1)/(d+1)} E(C_n) = 1. \tag{6}$$

2. Methods

Following Rényi and Sulanke [11], Efron [4], and Raynaud [10], we may express

$$E(F_n) = \binom{n}{d} W \tag{7}$$

where W is the probability that all points $X_{d+1}, X_{d+2}, \ldots, X_n$ lie on the same side of the (almost surely $(d-1)$-dimensional) hyperchord $\mathcal{H}$ through $X_1, X_2, \ldots, X_d$. Denote by R the distance from this chord to the center of S and by $L(R)$ the volume of the least of the two halves of S separated by $\mathcal{H}$. Denote by f_R the density of R. Raynaud showed $f_R(r)$ to be equal to $(1-r^2)^{(d^2-1)/2}$ up to a multiplicative constant ([10], Lemma 2).

Similarly, we may express

$$E(H_n) = \binom{n}{d} Y \tag{8}$$

where Y is the expected contribution of $\mathrm{Vol}(\mathrm{Conv}\{0, X_1, X_2, \ldots, X_d\})$ to the volume of A_n, provided $\{X_1, X_2, \ldots, X_d\}$ describe a facet. This contribution may be positive ($\{0, X_{d+1}, X_{d+2}, \ldots, X_n\}$ are on the same side of $\mathcal{H}$) or negative ($\{X_{d+1}, \ldots, X_n\}$ are on one side of $\mathcal{H}$ and 0 is on the other).

$$W = \int_0^1 \left[\left(1 - \frac{1}{v_d} L(r)\right)^{n-d} + \left(\frac{1}{v_d} L(r)\right)^{n-d} \right] f_R(r)\, dr \tag{9}$$

and

$$Y = \int_0^1 \left[\left(1 - \frac{1}{v_d} L(r)\right)^{n-d} - \left(\frac{1}{v_d} L(r)\right)^{n-d} \right] h(r) \cdot (r/d)\, dr \tag{10}$$

where $h(r)$ is obtained by integrating the product of the $(d-1)$-dimensional volume of $\mathrm{Conv}\{X_1, X_2, \ldots, X_d\}$ and the Jacobian that leads to $f_R(r)$, over all variables except r. This is so because the volume of $\mathrm{Conv}\{0, X_1, X_2, \ldots, X_d\}$ is $\mathrm{Vol}(\mathrm{Conv}\{X_1, X_2, \ldots, X_d\}) r/d$. $h(r)$ will be shown to be proportional to $(1-r^2)^{(d-1)(d+1)/2}$. The proportionality factor can be identified by the relation

$$\lim_{n \to \infty} E(H_n) = v_d. \tag{11}$$

$E(V_n)$ is hard to evaluate asymptotically because a second order term in $E(H_n)$ is required (see (1)). However, $E(T_n)$ and $E(C_n)$ are simple.

$$E(T_n) = nZ \tag{12}$$

and

$$E(C_n) = nU \tag{13}$$

where Z is the probability that $\mathrm{Conv}\{X_2, X_3, \ldots, X_n\}$ is disjoint to the hyperchord through X_1 that is perpendicular to X_1, and U is the probability that none of the points $X_2, X_3, \ldots, X_n$ belongs to the convex hull $\mathrm{Conv}(S \cup \{X_1\})$. Denote by $M(R)$ the volume of $\mathrm{Conv}(S \cup \{X_1\}) - S$ if $|X_1| = R$. Then,

$$Z = d\int_0^1 \left(1 - \frac{1}{v_d} L(r)\right)^{n-1} r^{d-1}\, dr \tag{14}$$

and

$$U = \frac{d}{(1+\Delta)^d - 1}\int_1^{1+\Delta} \left(1 - \frac{1}{v_d((1+\Delta)^d - 1)} M(r)\right)^{n-1} r^{d-1}\, dr. \tag{15}$$

3. Technical lemmas

In this section we will identify the functions defined in the previous section. We state for reference, omitting the proof,

LEMMA 1.

$$v_d = \pi^{d/2}/\Gamma\left(\frac{d}{2} + 1\right), \tag{16}$$

$$(1/\eta(m) =) \int_0^{\pi/2} \sin^m x\, dx = \frac{\Gamma\left(\frac{3}{2}\right)\Gamma\left(\frac{m}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{m}{2} + 1\right)}. \tag{17}$$

LEMMA 2.

$$L(r)/v_d = \frac{1}{2}\,\eta(d) \int_0^{\arccos r} \sin^d\theta\, d\theta. \tag{18}$$

PROOF. $L(r)$ may be computed as the integral over the hyperchord $\mathcal{H}$ at distance r from 0, of the distance (measured vertical to $\mathcal{H}$) from $\mathcal{H}$ to the surface of S. This distance, over a point in $\mathcal{H}$ a distance s from its center, is $\sqrt{1 - s^2} - r$, which is a function of s only. By Courant ([3], Chapter 4, Section 3) such a function may be integrated as follows:

$$\begin{aligned} L(r) &= \iint \cdots \int (\sqrt{1 - s^2} - r)^+\, dx_1\, dx_2 \cdots dx_{d-1} \\ &= \left(2\pi^{(d-1)/2}/\Gamma\left(\frac{d-1}{2}\right)\right) \int_0^{\sqrt{1-r^2}} (\sqrt{1 - s^2} - r) s^{d-2}\, ds \\ &= \left(\pi^{(d-1)/2}/\Gamma\left(\frac{d-1}{2}\right)\right) \int_0^{\arccos r} \sin^d\theta\, d\theta. \end{aligned}$$

■

LEMMA 3.

$$M(r)/v_d = (1/2)\eta(d)\left[(r/d)(1-(1/r^2))^{(d+1)/2} - \int_0^{\arccos(1/r)} \sin^d\theta\, d\theta\right]. \tag{19}$$

PROOF. Similar to the proof of Lemma 2. The hyperchord over which integration is performed goes through the tangency region of the tangents from the point outside the inner ball to the inner ball. The distance, measured vertically to this chord at a distance s from its center, from the surface of S to the tangent cone, is $r - s\sqrt{r^2-1} - \sqrt{1-s^2}$, so

$$M(r) = \left(2\pi^{(d-1)/2}/\Gamma\left(\frac{d-1}{2}\right)\right)\int_0^{\sqrt{1-(1/r^2)}} (r - s\sqrt{r^2-1} - \sqrt{1-s^2})s^{d-2}\, ds$$

$$= \left(\pi^{(d-1)/2}/\Gamma\left(\frac{d+1}{2}\right)\right)\left[(r/d)(1-(1/r^2))^{(d+1)/2} - \int_0^{\arccos(1/r)} \sin^d\theta\, d\theta\right].$$

■

LEMMA 4 (Raynaud [10], Lemma 2). *For* $0 < r < 1$,

$$f_R(r) = \eta(d^2)(1-r^2)^{(d^2-1)/2}. \tag{20}$$

LEMMA 5. *$h(r)$ is a constant multiple of* $(1-r^2)^{(d-1)(d+2)/2}$.

PROOF. The volume introduces a power $d-1$ of $\sqrt{1-r^2}$ to the integrand considered in the proof of Lemma 4. Hence the power of $\sqrt{1-r^2}$ must be $d-1+d^2-1=(d-1)(d+2)$. ■

LEMMA 6. *Let $g(x)$ be a bounded function on* $[0,1]$ *with* $\lim_{x\to 0} g(x) = 1$. *Let a, b and c be positive numbers and let f be a real number. Then*

$$\int_0^1 (1-g(x)cx^a)^{n+f}x^{b-1}\, dx = (1/a)c^{-(b/a)}n^{-(b/a)}\Gamma((b/a))\,[1+O((1/n))] \tag{21}$$

as $n \to \infty$.

PROOF. See any calculus book that studies the Gamma function (e.g., Franklin [5], article 328) or use directly $0 \le 1 - e^t(1-(t/n))^n \le (t^2/n)$. ■

4. Results

For completeness, we quote

THEOREM 1 (Raynaud [10], Théorème). *Denote*

$$\phi_m(t) = \eta(m)\int_0^t \sin^m\theta\, d\theta, \qquad 0 < t \le \pi/2. \tag{22}$$

Then

$$(23)\qquad E(F_n) = \binom{n}{d}\int_0^{\pi/2} [(1-\phi_d(t)/2)^{n-d} + (\phi_d(t)/2)^{n-d}]\, d\phi_{d^2}(t),$$

$$(24)\qquad \begin{aligned}&\lim_{n\to\infty} n^{-(d-1)/(d+1)}E(F_n)\\ &\quad = (2(d+1)/\eta(d))^{(d^2+1)/(d+1)}\left(\Gamma\left(\frac{d^2+1}{d+1}\right)\Big/(d+1)!\right)\eta(d^2).\end{aligned}$$

PROOF. Expression (23) is obtained by substituting (18) and (20) in (9). To obtain expression (24), apply (21) after expanding $L(r)$ through $r=1$ to obtain that

$$(25)\qquad L(r)/v_d = g(1-r)(1-r)^{(d+1)/2}2^{(d-1)/2}\eta(d)/(d+1),$$

where $g(x)\to 1$ as $x\to 0$. $(1-r^2)$ is replaced by $2(1-r)$. ■

To approximate (24) for large d, use the following two approximations:

$$(26)\qquad \lim_{m\to\infty}\frac{\Gamma(m+a/m)}{\Gamma(m)} = \lim_{m\to\infty}\left(\frac{\Gamma(m+1)}{\sqrt{m-(1/4)}\,\Gamma(m+(1/2))}\right)^m = 1.$$

The proofs of Theorems 2 and 3 are omitted.

THEOREM 2. *Let F be the uniform distribution on $S_{1+\Delta}^{(d)} - S_1^{(d)}$, where $S_r^{(m)}$ is the m-dimensional ball of radius r with center at zero. Then, if $X_1, X_2, \ldots, X_n$ are independent and identically F-distributed,*

$$(27)\qquad E(C_n) = \frac{nd}{(1+\Delta)^d - 1}\int_1^{1+\Delta}\left(1 - \frac{1}{v_d((1+\Delta)^d-1)}M(r)\right)^{n-1} r^{d-1}\, dr,$$

$$(28)\qquad \begin{aligned}&\lim_{n\to\infty} n^{-(d-1)/(d+1)}E(C_n)\\ &\quad = \frac{d}{2((1+\Delta)^d-1)}(\pi d\eta(d+1)((1+\Delta)^d-1))^{2/(d+1)}\Gamma\left(\frac{d+3}{d+1}\right).\end{aligned}$$

Let $a_d = d\Delta$. If

$$\liminf_{d\to\infty} a_d > 0 \quad \textit{and} \quad \lim_{d\to\infty} a_d^2/d = 0$$

then

$$(29)\qquad \lim_{d\to\infty}(\lim_{n\to\infty} n^{-(d-1)/(d+1)}E(C_n))^{-1}\frac{d}{2(e^{a_d}-1)} = 1.$$

In particular, if $\Delta = 2^{(1/d)} - 1$, *i.e., if the inner ball has half the volume of the bigger ball,*

$$\lim_{d\to\infty}(\lim_{n\to\infty} n^{-(d-1)/(d+1)}E(C_n))^{-1}(d/2) = 1. \tag{30}$$

THEOREM 3.

$$E(T_n) = nd\int_0^1 \left(1 - \frac{1}{v_d}L(r)\right)^{n-1} r^{d-1}\,dr, \tag{31}$$

$$\lim_{n\to\infty} n^{-(d-1)/(d+1)}E(T_n) = (\pi\eta(d+1))^{2/(d+1)}(d/2)\Gamma\left(\frac{d+3}{d+1}\right), \tag{32}$$

$$\lim_{d\to\infty}(\lim_{n\to\infty} n^{-(d-1)/(d+1)}E(T_n))^{-1}(d/2) = 1. \tag{2}$$

THEOREM 4.

$$\begin{aligned}\frac{E(H_n)}{v_d} &= \binom{n}{d} d(d+1)(\pi\eta(d+1))^{-d}\\ &\times\int_0^1 \left[\left(1-\frac{1}{v_d}L(r)\right)^{n-d} - \left(\frac{1}{v_d}L(r)\right)^{n-d}\right] r(1-r^2)^{(d+2)(d-1)/2}\,dr,\end{aligned} \tag{33}$$

$$\frac{E(H_{d+1})}{v_d} = \pi^{-d}\frac{\eta((d+1)^2)}{(\eta(d+1))^{d+1}}, \tag{3}$$

$$E(V_n) = n[1 - E(H_{n-1})/v_d], \tag{1}$$

$$\lim_{n\to\infty} n^{-(d-1)/(d+1)}E(V_n) = \frac{d(d+1)}{2(d+3)}\,\frac{\Gamma(d+1+2/(d+1))}{\Gamma(d+1)}\,(\pi\eta(d+1))^{2/(d+1)}, \tag{34}$$

$$\lim_{d\to\infty}(\lim_{n\to\infty} n^{-(d-1)/(d+1)}E(V_n))^{-1}(d/2) = 1. \tag{2}$$

PROOF. By Lemma 5 and expression (10), (33) has been proved up to the identification of the multiplicative constant; let $n \to \infty$, use Lemma 6 and identify the limit as v_d.

Expression (3) follows by straightforward integration of (33) with $n = d + 1$. Expression (1) is due to the fact that X_1 is a vertex if and only if it does not belong to the convex hull of $\{X_2, X_3, \ldots, X_n\}$. As $n \to \infty$, we may disregard in expression (33) the negative term, since it contributes an order of magnitude $n^{d+1}2^{-n}$ to the expression. For similar reasons, we may disregard the difference between $\int_0^{1/2}(1-y)^{n-d}\psi(y)\,dy$ and $\int_0^1(1-y)^{n-d}\psi(y)\,dy$ for polynomial ψ. Instead of obtaining a second term directly, we will express the 1 in (1) as an integral very close to (33) and estimate the difference of the integrands.

Letting

$$y = \frac{1}{v_d} L(r), \qquad dy = (1/2)\eta(d)(1 - r^2)^{(d-1)/2}\, dr,$$

(33) applied to (1) yields

$$1 - \frac{E(H_n)}{v_d} = O(1/n) + n^d \frac{1}{(d-1)!} \int_0^1 (1/y)^{n-d}$$
$$\times \left[y^{d-1} - \left(\frac{\eta(d)}{2(d+1)} \right)^{d-1} r(1 - r^2)^{(d^2-1)/2} \right] dy.$$

The difference in brackets is asymptotic (as $r \to 1$) to

$$(d-1)\left(y - \frac{\eta(d)}{2(d+1)} r^{1/(d-1)} (1 - r^2)^{(d+1)/2} \right) y^{d-2},$$

which in turn is asymptotic to

$$((d^2 + d + 2)/(2(d+3)))(2(d+1)/\eta(d))^{2/(d+1)} y^{d-1+2/(d+1)}$$

and, finally, Lemma 6 yields (34). ∎

We will compute the asymptotic ratio of $E(V_n)$ to $E(T_n)$:

$$\lim_{n\to\infty} \frac{E(V_n)}{E(T_n)} = \frac{\Gamma(d+1+2/(d+1))}{\Gamma(d+1)\Gamma(2+2/(d+1))} \tag{35}$$

This expression equals 1 at $d = 1$, converges to 1 as $d \to \infty$ and is strictly bigger than 1 for $d \geq 2$, since $\Gamma(m+\alpha)/\Gamma(m) = ((m+\alpha)/m)\Gamma(m-1+\alpha)/\Gamma(m-1) > \Gamma(m-1+\alpha)/\Gamma(m-1)$, so $\Gamma(m+\alpha)/\Gamma(m)$ increases with m. Hence, $\Gamma(m+\alpha)/\Gamma(m) > \Gamma(2+\alpha)/\Gamma(2) = \Gamma(2+\alpha)$.

Replace $m = d+1$, $\alpha = 2/(d+1)$. Ratio (35) is maximal at $d = 5$ and equals 1.498. As for the ratio of $E(C_n)$ to $E(V_n)$ for $\Delta = 2^{(1/d)} - 1$,

$$\lim_{n\to\infty} \frac{E(C_n)}{E(V_n)} = d^{2/(d+1)}. \tag{36}$$

This ratio is maximal at $d = 4$, where it equals 1.741.

5. A statistical application

A random sample of n points is taken from a known continuous distribution F on R^d that assigns (an unknown) probability α to an unknown compact convex set S. We wish to estimate the set S and the probability $\alpha = F(S)$ on the basis of the random sample $X_1, X_2, \ldots, X_n$ and the indicators $\psi_1, \psi_2, \ldots, \psi_n$ of the events "$X_i \in S$."

Let A_n be the convex hull of the sample points in S and let B_n be *shade* of A_n beyond the sample points outside S, i.e., all points outside S that can be connected to some point in A_n by a linear segment that goes through some sample point outside S. Mathematically,

$$(37) \quad B_n = \bigcup_{\{i \mid \psi_i = 0\}} \left\{ z \,\middle|\, z = X_i + \lambda(X_i - y) \text{ for some } \lambda \geq 0 \text{ and some } y \in A_n \right\}.$$

Let $\mathfrak{V}_n$ be the set of, and V_n the number of, vertices in A_n and let $\mathfrak{C}_n$ be the set of, and C_n the number of, *cusps* in B_n, where a cusp is a sample point outside S whose removal would change the set B_n.

We shall show "the picture" $(\mathfrak{V}_n, \mathfrak{C}_n)$ to be a minimal sufficient statistic for S. We shall also sketch a proof that if $\hat{\hat{\alpha}}$ is the proportion of sample points in S, then the unbiased Rao–Blackwell improvement of $\hat{\hat{\alpha}}$, $\hat{\alpha} = E_S\left(\hat{\hat{\alpha}} \,\middle|\, (\mathfrak{V}_n, \mathfrak{C}_n)\right)$, is

$$(38) \qquad \hat{\alpha} = \frac{V_n}{n} + \frac{n - V_n - C_n}{n} \; \frac{F(A_n)}{F(A_n) + F(B_n)}$$

whose asymptotic variance as $n \to \infty$ is

$$(39) \qquad \operatorname{Var}(\hat{\alpha}) \sim \frac{\alpha(1-\alpha)}{n} \left[\frac{1-\alpha}{\alpha} E\left(\frac{V_n}{n}\right) + \frac{\alpha}{1-\alpha} E\left(\frac{C_n}{n}\right) \right].$$

Since the variance of $\hat{\hat{\alpha}}$ is $\alpha(1-\alpha)/n$, the convexity of S admits an unbiased estimate of $F(S)$ infinitely more efficient than the sample proportion. We essentially showed in Theorems 2 and 4 that for smooth S and F, $\operatorname{Var}(\hat{\alpha})/\operatorname{Var}(\hat{\hat{\alpha}})$ is asymptotic to $n^{(d-1)/(d+1)}$ as $n \to \infty$.

The observations $(X_1, \psi_1), (X_2, \psi_2), \ldots, (X_n, \psi_n)$, are independent and identically distributed, with a distribution that belongs to the family $\mathfrak{F} = \{\tilde{F}_S \mid S \in \mathcal{S}\}$, where $\mathcal{S}$ is the collection of compact convex sets in R^d and

$$(40) \qquad \tilde{F}_S(X \in A,\ \psi = 1) = F(X \in A \cap S), \qquad \tilde{F}_S(X \in A) = F(X \in A).$$

CLAIM. The pair $(\mathfrak{V}_n, \mathfrak{C}_n)$ is a minimal sufficient statistic for the family $\mathfrak{F}$.

PROOF. The likelihood function is

$$\prod_{i=1}^{n} d\tilde{F}_S(x_i, \psi_i) = \chi(S, A_n, B_n) \prod_{i=1}^{n} dF(x_i),$$

where $\chi(S, A_n, B_n)$ is the indicator function of the event "S contains A_n and is disjoint from B_n." Hence, $(\mathfrak{V}_n, \mathfrak{C}_n)$ (or, equivalently, (A_n, B_n)) is sufficient. It is also minimal, since (A_n, B_n) can be reconstructed from the values of $\chi(S, A_n, B_n)$ for all S as follows: A_n is the intersection and the complement of B_n is the union of all S for which $\chi(S, A_n, B_n) = 1$. ■

Proof of (38). Given $(\mathfrak{B}_n, \mathfrak{C}_n)$, the number of sample points that belong to S among the $n - V_n - C_n$ that are neither vertices nor cusps, is binomial with probability $F(A_n)/(F(A_n) + F(B_n))$ of success. ■

Sketch of the Proof of (39).

$$\begin{aligned}
\mathrm{Var}_S(\hat{\alpha}) &= \mathrm{Var}_S E_S(\hat{\hat{\alpha}} \mid \mathfrak{B}_n, \mathfrak{C}_n) = \mathrm{Var}_S(\hat{\hat{\alpha}}) - E_S(\mathrm{Var}_S(\hat{\hat{\alpha}} \mid \mathfrak{B}_n, \mathfrak{C}_n)) \\
&= \frac{\alpha(1-\alpha)}{n} - E_S\left[\frac{n - V_n - C_n}{n^2} \, \frac{F(A_n) * F(B_n)}{(F(A_n) + F(B_n))^2}\right] \\
&= \frac{\alpha(1-\alpha)}{n} - \frac{1}{n} E_S\left[\left(\hat{\alpha} - \frac{V_n}{n}\right)\left(1 - \frac{n}{n - V_n - C_n}\left(\hat{\alpha} - \frac{V_n}{n}\right)\right)\right].
\end{aligned}$$

Let us express $\hat{\alpha} = \alpha + Z\sqrt{\mathrm{Var}_S(\hat{\alpha})}$. Then

$$\begin{aligned}
\mathrm{Var}_S(\hat{\alpha}) = \frac{\alpha(1-\alpha)}{n} - \frac{1}{n} E_S\Bigg[&\left(\alpha + Z\sqrt{\mathrm{Var}_S(\hat{\alpha})} - \frac{V_n}{n}\right) \\
&\times \left(1 - \frac{n}{n - V_n - C_n}\left(\alpha + Z\sqrt{\mathrm{Var}_S(\hat{\alpha})} - \frac{V_n}{n}\right)\right)\Bigg] \\
= \frac{1}{n} E_S\Bigg[&\mathrm{Var}_S(\hat{\alpha}) + \frac{V_n}{n}(1 - 2\alpha) - \frac{2ZV_n}{n}\sqrt{\mathrm{Var}_S(\hat{\alpha})} \\
&+ \frac{V_n^2}{n^2} + \frac{V_n + C_n}{n - V_n - C_n}\left(\alpha + Z\sqrt{\mathrm{Var}_S(\hat{\alpha})} - \frac{V_n}{n}\right)^2\Bigg].
\end{aligned}$$

Since $\mathrm{Var}(\hat{\alpha}) \le \mathrm{Var}(\hat{\hat{\alpha}}) \to 0$ and V_n/n and C_n/n are bounded and converge to zero a.s. as $n \to \infty$, some terms are negligible with respect to others, and we obtain

$$\begin{aligned}
\mathrm{Var}_S(\hat{\alpha}) &\sim (1 - 2\alpha) E_S\left(\frac{V_n}{n^2}\right) + \alpha^2 E_S\left(\frac{V_n + C_n}{n^2}\right) \\
&= (1 - \alpha)^2 E_S\left(\frac{V_n}{n^2}\right) + \alpha^2 E_S\left(\frac{C_n}{n^2}\right).
\end{aligned}$$ ■

Acknowledgements

This research was carried out while the author visited IBM T.J. Watson Research Center in Yorktown Heights, N.Y. Thanks are due to Gary Hachtel for initiating the subject, to Arthur Nádas for the simulations and computations, to both for the joint work, and to Don Coppersmith for the idea of using (11).

References

1. I. Bárány, *Intrinsic volumes and f-vectors of random polytopes*, Math. Ann. **285** (1989), 671-699.

2. C. Buchta, *Zufällige Polyeder—Eine Übersicht, Zahlentheoretische Analysis*, Lecture Notes in Mathematics **1114**, Springer-Verlag, Berlin, 1985, pp. 1-13.

3. R. Courant, *Differential and Integral Calculus*, Vol. 2, Interscience Publishers, Inc., New York, 1936.

4. B. Efron, *The convex hull of a random set of points*, Biometrika **52** (1965), 331-343.

5. P. Franklin, *A Treatise on Advanced Calculus*, John Wiley & Sons, Inc., New York, 1940.

6. P. Groeneboom, *Limit theorems for convex hulls*, Probab. Theory Relat. Fields **79** (1988), 327-368.

7. B. Grünbaum, *Convex Polytopes*, John Wiley & Sons, Inc., New York, 1967.

8. G. Hachtel, I. Meilijson and A. Nádas, *The estimation of a convex set in R^d and its probability content*, RC 8666 (#37890), IBM T.J. Watson Research Center, 1981.

9. H. Raynaud, *Sur le comportement asymptotique d l'enveloppe convexe d'un nuage tirés au hasard dans R^n*, C.R. Acad. Sci. Paris, **261** (19 juillet 1965), Groupe 1.

10. H. Raynaud, *Sur l'enveloppe convexe des nuages de points aleatoires dans R^n*. I, J. Appl. Prob. **7** (1970), 35-48.

11. A. Rényi and R. Sulanke, *Uber die konvexe Hülle von n zufällig gewählten Punkten*, Z. Wahrschenlichkeitstheor. Verw. Geb., Part I: **2** (1963), 75-84; Part II: **3** (1964), 138-147.

12. R. Schneider and J.A. Wieacker, *Random polytopes in a convex body*, Z. Wahrscheinlichkeitstheor. Verw. Geb. **52** (1980), 69-73.