On the Variance of Random Polygons

Imre Bárány* and William Steiger†

Abstract

A random polygon is the convex hull of uniformly distributed random points in a convex body $K \subset \mathbf{R}^2$. General upper bounds are established for the variance of the area of a random polygon and also for the variance of its number of vertices. The upper bounds have the same order of magnitude as the known lower bounds on variance for these functionals. The results imply a strong law of large numbers for the area and number of vertices of random polygons for all planar convex bodies. Similar results had been known, but only in the special cases when K is a polygon or where K is a smooth convex body. The careful, technical arguments we needed may lead to tools for analogous extensions to general convex bodies in higher dimension. On the other hand one of the main results is a stronger version in dimension $d = 2$ of the *economic cap covering theorem* of Bárány and Larman. It is crucial to our proof, but it does not extend to higher dimension.

1 The main results

Let $K \subset \mathbf{R}^d$ be a convex set of volume one (we write $V(K) = 1$) and let $x_1, \dots, x_n$ be a random sample of n independent, identically distributed points chosen uniformly from K. The *random polytope* $K_n \equiv \mathrm{Conv}\{x_1, \dots, x_n\}$ is the convex hull of these points. Understanding the asymptotic behaviour of K_n is one of the classical problems in stochastic geometry. Starting with Rényi and Sulanke [7] in 1963, there have been many results concerning the expectation of various functionals of K_n. For instance the expectation of the random variables like the missed volume $V(K \setminus K_n)$, and of $f_0(K_n)$, the number vertices of K_n, have been determined with high precision (see e.g., the book by Schneider and Weil [10].

Determining the variance is more difficult. For *smooth* convex bodies its order of magnitude was determined by Reitzner [8] and [9]. Schreiber and Yukich [11] have computed the precise asymptotic behaviour of the variance of $f_0(K_n)$ when K is the unit ball, a significant breakthrough. Recently Bárány and Reitzner [3] obtained a lower bound on the the variance of $V(K_n)$ and also of $f_\ell(K_n)$ for general convex bodies (f_ℓ counts the number of ℓ-dimensional faces).

In order to state the results we need a few definitions. First, $v : K \to \mathbf{R}$ is the function given by

$$v(z) = \min\{V(K \cap H) : \ H \text{ is a halfspace and } z \in H\}.$$

The *floating body* with parameter t is just the level set $K(v \geq t) = \{z \in K : \ v(z) \geq t\}$, which is clearly convex. The set $K(v \leq t)$ is called the *wet part*, that is, where v is at most t. The general lower bound for variance is

Proposition 1 *([3]) Assume $K \subset \mathbf{R}^d$ is a convex body of volume one. Then*

$$\begin{aligned} n^{-1} V(K(v \leq n^{-1})) &\ll \mathrm{Var} V(K \setminus K_n) \\ n V(K(v \leq n^{-1})) &\ll \mathrm{Var} f_\ell(K_n) \end{aligned}$$

We use Vinogradov's $f(n) \ll g(n)$ notation which means that there are constants n_0 and $c_0 > 0$ (depending possibly on d but not on K) such that $f(n) \leq c_0 g(n)$ for every $n \geq n_0$.

The main contribution of the present paper is a matching upper bound for the planar case $d = 2$.[1]

Theorem 2 *Assume $K \subset \mathbf{R}^2$ is a convex body of area one. Then*

$$\begin{aligned} \mathrm{Var} V(K \setminus K_n) &\ll n^{-1} V(K(v \leq n^{-1})) \\ \mathrm{Var} f_0(K_n) &\ll n V(K(v \leq n^{-1})). \end{aligned}$$

Note that the constants implied by the $\ll$ notation are universal because $d = 2$. An advantage of this kind of result is that it is usually much easier to compute the volume of the wet part than the variance of K_n.

Statements similar to Theorems 1 and 2 are known for the expectations (see [2]), for instance

$$V(K(v \leq n^{-1})) \ll \mathbb{E} V(K \setminus K_n) \ll V(K(v \leq n^{-1})).$$

This fact and Theorem 2 combine to imply a strong law of large numbers for $V(K_n)$ and for $f_0(K_n)$ in the plane.

*Rényi Institute, Hungarian Academy of Sciences and Department of Mathematics, University College, London

†Department of Computer Science, Rutgers University, 110 Frelinghuysen Road, Piscataway, New Jersey 08854-8004

[1] This result is the first nontrivial case of the conjecture from [3] that states the same upper bounds in all dimensions. It was known to be true for smooth convex bodies (see Reitzner [9]), and a slightly weaker upper bound is proved in [3] when K is a polytope. Recently John Pardon [6] has proved the same upper bound, and much more.

Corollary 3 *Assume $K \subset \mathbf{R}^2$ is a convex body of area one and let K_n be the random polygon generated by a uniform sample of n points from K. Then*

$$Prob\{n^{2/3}V(K \setminus K_n) \to c_1\} = 1,$$

$$Prob\{n^{-1/3}f_0(K_n) \to c_2\} = 1,$$

where c_1 and c_2 are constants depending on K.

Reitzner [8] obtained similar statements for the case of smooth convex bodies by appealing to Tchebycheff, the Borel-Cantelli lemma, and an argument about convergence of subsequences. We can prove it directly using the complete convergence theorem of Hsu and Robbins [5]

Theorem 2 is a direct consequence of a strengthened version of the economic cap covering theorem of Bárány and Bárány and Larman that holds in dimension 2, and is of independent interest. Specifically we prove

Theorem 4 *Let $K \subset \mathbf{R}^2$ be a convex body of area 1. There are numbers $T_0 > 0$ and $q \in (0,1)$ such that for all $T \in (0, T_0]$ and for all $t \in (0, qT]$ the following holds. For every cap D of K of area T and for every cap covering $C_1, C_2, \ldots, C_m$ of $K(v \leq t)$,*

$$V(K(v \leq t) \cap D) \ll \sum_{i=1}^{m} V(C_i \cap D) \ll V(K(v \leq t) \cap D).$$

In the next section we will review cap covering and explain how Theorem 4 implies the truth of a conjecture of Bárány and Reitzner [3] which already had been shown to imply Theorem 2. Finally, in the remaining sections, we will sketch the proof of Theorem 4 and thereby, of Theorem 2.

2 Economic cap coverings

We fix the convex body $K \subset \mathbf{R}^d$ of volume one. A cap C of K is the intersection of K with a closed halfspace H. The *centre* of C is a point $x \in C$ (not necessarily unique) with maximal distance from the bounding hyperplane, L, of H. The *width* of C, $w(C)$, is just the distance between x and L. For $\lambda > 0$ let H^λ be the halfspace containing H for which the width of the cap $C^\lambda = K \cap H^\lambda$ is λ times the width of C. Observe that for $\lambda \geq 1$, $C^\lambda \subset x + \lambda(C - x)$, implying that $V(C^\lambda) \leq \lambda^d V(C)$ if $\lambda \geq 1$.

The *minimal cap* of $z \in K$ is a cap $C(z)$ containing z such that $v(z) = V(C(z))$. Again, it need not be unique.

The *Macbeath region, or M-region,* for short, with center z and factor $\lambda > 0$ is

$$M(z, \lambda) = M_K(z, \lambda) = z + \lambda[(K - z) \cap (z - K)].$$

The M-region with $\lambda = 1$ is just the intersection of K and K reflected with respect to z. Thus $M(z, 1)$ is convex and centrally symmetric with cente z, and $M(z, \lambda)$ is a homothetic copy of $M(z, 1)$ with center z and factor of homothety λ. The following lemma, originally from [4], is crucial.

Lemma 5 *Suppose $M(x, 1/2) \cap M(y, 1/2) \neq \emptyset$. Then $M(x, 1) \subset M(y, 5)$.*

Set

$$t_0 = (16d)^{-2d}. \quad (1)$$

The boundary of $K(v \geq t)$ is clearly $K(v = t)$. Assume $t \leq t_0$ and choose a maximal system of points $X = \{x_1, \ldots, x_m\}$ on $K(v = t)$ having pairwise disjoint M-regions $M(x_i, 1/2)$. Such a system will be called *saturated*. Note that X (and even m) is not defined uniquely. Clearly $V(C(x_i)) = t$. Set

$$K_i = M(x_i, 1/2) \cap C(x_i) \text{ and } C_i = (C(x_i))^{16}.$$

We write $[m]$ for $\{1, 2, \ldots, m\}$. The following result, the so called *economic cap covering theorem*, comes from Theorem 6 in [2] and Theorem 7 in [1]. The present form is copied here from [3].

Proposition 6 *Suppose $t \in (0, t_0]$, $K \subset \mathbf{R}^d$ is a convex body of volume one, and $X = \{x_1, \ldots, x_m\}$ is a saturated system on $K(v = t)$. Then, with C_i and K_i as defined above, the following holds*

(i) $\bigcup_1^m K_i \subset K(v \leq t) \subset \bigcup_1^m C_i$,

(ii) $t \leq V(C_i) \leq 16^d t$, *for* $i \in [m]$,

(iii) $(6d)^{-d} t \leq V(K_i) \leq 2^{-d} t$, $i \in [m]$,

(iv) every C with $V(C) \leq t$ is contained in some C_i with $i \in [m]$.

The sets $C_1, \ldots, C_m$ from this construction will be called an *economic cap covering* of $K(v \leq t)$.

The following conjecture was stated in [3].

Conjecture 7 *For every $d \geq 2$ there are numbers $T_0 > 0$ and $q \in (0,1)$ such that for all convex bodies $K \subset \mathbf{R}^d$ of volume one, and for all $T \in (0, T_0]$, and for all $t \in (0, qT]$ the following holds. Let $D_1, \ldots, D_{m(T)}$, resp. $C_1, \ldots, C_{m(t)}$ be the covering caps for $K(v \leq T)$ and $K(v \leq t)$ from Theorem 6. Then*

$$\sum_{i=1}^{m(T)} V(K(v \leq t) \cap D_i) \ll \sum_{i=1}^{m(T)} \sum_{j=1}^{m(t)} V(C_j \cap D_i)$$

$$\ll \sum_{i=1}^{m(T)} V(K(v \leq t) \cap D_i).$$

It was proved in [3] that this conjecture would imply the general upper bound, of the same order as in the lower bounds in Theorem 1, on the variances of the random variables $V(K \setminus K_n)$ and $f_\ell(K_n)$. To see that the conjecture is true for $d = 2$, simply apply the inequalities of Theorem 4 to each D_i and sum the results. The left hand side inequality is a direct consequence of the cap covering theorem.

Theorem 4 is a strengthening, in dimension 2, of the above conjecture. Unfortunately the theorem does not remain true in higher dimensions. Suppose $K \subset \mathbf{R}^d$ ($d \geq 3$) is the truncated cone $x_1^2 + \dots x_{d-1}^2 \leq x_d \leq 1$. The cap D is cut off from K by the hyperplane $x_d \leq h$ for small $h > 0$ and the cap coverings of $K(v \leq t)$ go with small $t > 0$. A direct computation shows that $\sum_1^m V(C_i \cap D)$ is not smaller than a constant times $V(K(v \leq t) \cap D)$; we omit the details.

In the next section we prepare some facts needed for the proof of Theorem 4, and then we sketch the proof.

3 Auxiliary Lemmas and Preparations

Since $d = 2$ we will use Area instead of V. We briefly review some properties of the M-regions and minimal caps. We assume that $t \leq t_0$ where $t_0 = 32^{-4}$, but certainly t_0 could be taken much larger. We make no effort to minimize constants.

The floating body $K(v \geq t)$ is convex. It was shown in [1] that its boundary $K(v = t)$, contains no line segment. This implies that if C is a cap with Area $C \leq t_0$, then $\max\{v(x) : x \in C\}$ is reached at a unique point $z \in C$; actually, with $v(z) = t$, $C \cap K(v \geq t) = \{z\}$. So z lies on the bounding segment, $[a, b]$ of C. The convex curve $K(v = t)$ has unique left and right tangents at z that cut off caps C^{left} and C^{right} from K. As was shown in [1]

Lemma 8 *$t =$ Area $C^{left} =$ Area C^{right} and $C \subset C^{left} \cup C^{right}$. In particular, $t \leq$ Area $C \leq 2t$.*

If the left and right tangents to $K(v = t)$ at z coincide, then C is the minimal cap of z, and z is the midpoint of the bounding segment $[a, b]$ of C.

Lemma 9 *With the above notation $|a - z| \leq 2|b - z|$.*

The function $u : K \to \mathbf{R}$ is defined by $u(x) =$ Area $M(x, 1)$. Many things are known about $u(x)$. In particular it is shown in [2] that $u(x)$ and $v(x)$ are very close to each other near the boundary of K:

Lemma 10 *For every $x \in K$, $u(x) \leq 2v(x)$. If $v(x) \leq t_0$ or if $u(x) \leq t_0$, then $v(x) \leq 16u(x)$.*

We place the coordinate system so that $[b_1, b_2]$, the bounding segment of D, lies on the x-axis, and the origin

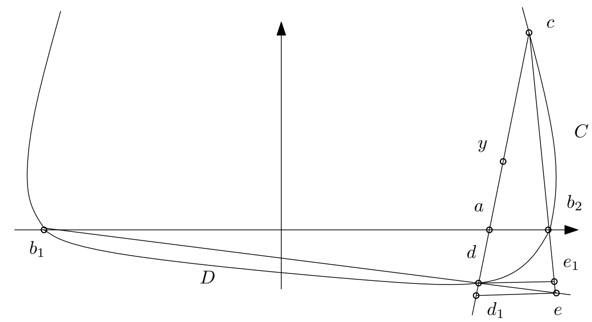

Figure 1: The caps D, C and the quadrilateral Q.

is the point where $v(x)$ takes its maximal value on D (see Figure 1). Lemma 8 shows that $v(0) \leq T \leq 2v(0)$.

In the next statement, b denotes either b_1 or b_2.

Lemma 11 *If $a \in [0, b]$ and $v(a) \leq 2^{-6}5^{-2k}T$, then $|b - a| \leq 3^{-k}|b|$.*

To see this, observe that Area $D = T \leq 2v(0)$. Now fix the constants T_0 and q: $T_0 = t_0 = 32^{-4}$ and suppose from now on that $T \leq T_0$. There will be an intermediate t^* satisfying

$$t^* = 2^9 t \text{ and } t^* \leq 2^{-6}5^{-4}T \tag{2}$$

so $q = 2^{-15}5^{-4}$.

Next let C be a cap with bounding segment $[c, d]$. Denote by y the point on $[c, d]$ where $v(x)$ reaches its maximal value on $x \in C$. Assume $[c, d]$ intersects $[b_1, b_2]$ in a point $a \in [0, b]$ where, again, b denotes either one of the points b_1 or b_2. We use the notation of Figure 1 (where $b = b_2$). The figure is distorted since Area C should be much smaller than Area D. We write Q for the quadrilateral with vertices a, b, e, d_1.

Lemma 12 *If $|b| > 3|a - b|$, Area $Q \ll$ Area $C \cap D \leq$ Area Q.*

Proof. The upper bound is trivial since $C \cap D \subset Q$. For the lower bound, let $h(x)$ denote the distance of $x \in \mathbf{R}^2$ from the x axis and let k be the smallest integer with $|b| \leq 3^k|b - a|$. Then $k \geq 2$ and $|b| > 3^{k-1}|a - b|$. Lemma 9 shows that $|y - d| \leq 2|y - c|$ implying $h(d) \leq 2h(c)$, and $|d - e_1| \leq 3|a - b|$. Then $|d - e_1|/|b_2 - b_1| = (h(e) - h(d))/h(e)$, implying

$$\frac{h(d)}{h(e)} \geq 1 - \frac{3|a - b|}{|b_2 - b_1|} \geq \beta,$$

where $\beta = 1 - 2 \cdot 3^{-k+1} > 0$; this follows from $|a - b| < 3^{-k+1}|b|$ and from $|b_2 - b_1| = |b_2| + |b_1| \geq |b| + |b|/2$ with Lemma 9. Now we have

$$\text{Area } Q = \left[\left(\frac{h(e) + h(c)}{h(c)} \right)^2 - 1 \right] \text{Area } [c, a, b_2]$$

$$= \frac{h(e)}{h(c)} \left(\frac{h(e)}{h(c)} + 2 \right) \frac{1}{2} h(c)|a - b_2|$$

$$= \frac{h(e)}{h(d)} \left(\frac{h(e)}{h(c)} + 2 \right) \frac{1}{2} h(d)|a - b_2|,$$

so Area $Q \leq (1/\beta)(2/\beta + 2)$Area $[a, b, d] \leq (1/\beta)(2/\beta + 2)$Area $C \cap D$. Note that β is increasing with k and $\beta \geq 1/3$ for all $k \geq 2$. Thus Area $Q \leq 24$Area $C \cap D$. ■

Remark. The lemma holds even if $a = y$, that is, when $v(x)$ reaches its maximal value on C at $x = a$. We will use it in this form in Lemma 16.

4 The proof, first part

We define $I_0 = \{i \in [m] : x_i \in D\}$.

Lemma 13 $\sum_{i \in I_0}$ Area $C_i \cap D \ll$ Area $K(v \leq t) \cap D$.

Proof. We may take $I_0 \neq \emptyset$ or there is nothing to prove. Using Area $C_i \leq 2t$ from Lemma 8, $\sum_{i \in I_0}$ Area $C_i \cap D \leq \sum_{i \in I_0}$ Area $C_i \leq 2t|I_0| \ll \sum_{i \in I_0}$ Area $M(x_i, 1/2)$ where the last inequality holds since $I_0 \neq \emptyset$ and since Area $M(x_i, 1/2) = \frac{1}{4}u(x_i) \leq \frac{1}{64}v(x_i) = t/2$ by Lemma 10. Further, $\sum_{i \in I_0}$ Area $M(x_i, 1/2) = 2\sum_{i \in I_0}$ Area K_i.

It is easy to show the following:

Claim 14 *When $i \in I_0$, $K_i \subset K(v \leq t) \cap D$ except possibly for the leftmost and rightmost K_i.*

Using now Area $K(v \leq t) \cap D \geq t$ we have $\sum_{i \in I_0}$ Area $K_i \leq$ Area $K(v \leq t) \cap D + 2t \ll$ Area $K(v \leq t) \cap D$. ■

Remark. This argument actually holds in any dimension.

For each $x_i \notin D$ we define the cap C_i^* whose bounding segment is parallel with that of C_i so that $C_i^* \cap K(v \geq t^*)$ is a single point y_i. Here t^* is given by $2^9 t = t^*$, according to (2). We claim that $C_i \subset C_i^*$ for every $i \in [m] \setminus I_0$. Indeed, Area $C_i \leq 16^2 t$ because $C_i = C(x_i)^{16}$. So even if C_i is not a minimal cap, it is disjoint from $K(v \geq t^*)$ as shown by Lemma 8. It is also clear that Area $C_i^* \ll t$. We are going to show that

$$\sum_{i \in [m] \setminus I_0} \text{Area } C_i^* \cap D \ll t, \qquad (3)$$

which will finish the proof since $C_i \subset C_i^*$.

Remark. This inequality does not hold for the example given at the end of Section 2.

Define $I_1 = \{i \in [m] : x_i \notin D \text{ and } y_i \in D\}$ and $I = [m] \setminus (I_0 \cap I_1)$. We will show that the contribution of the terms Area $C_i^* \cap D$ with $i \in I_1$ is not too large.

Lemma 15 $\sum_{i \in I_1}$ Area $(C_i^* \cap D) \ll t$.

This **proof** is simpler than the previous one. The wet part $K(v \leq t^*)$ intersects $[b_1, b_2]$ in two segments. Consider one of them, $[a, b_2]$ say. Let I^* be the set of those $i \in I_1$ for which the bounding segment of C_i intersects $[a, b_2]$. By symmetry it is enough to show that $\sum_{i \in I^*}$ Area $C_j^* \cap D \ll t$.

Let $j \in I^*$ be the element for which $h(y_j)$ is smallest. Then all other x_i with $i \in I^*$ lie in $C_j^* \setminus D \subset C_j^*$, and the corresponding K_i are pairwise disjoint, and all of them (except possibly the leftmost) are contained in C_j^*, and $K_j \subset C_j^*$, of course. Thus $|I^*| \ll \frac{1}{t}$Area $C_j^* + 1 \ll 1$ and this implies the lemma since Area $C_i^* \ll t$. ■

The final steps in the proof of Theorem 4 bound $\sum_{i \in I}$ Area $C_i^* \cap D$,; $I = [m] \setminus (I_0 \cap I_1)$. This is more difficult, and for reasons of space, we have to postpone these arguments for the full paper.

References

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

[2] Bárány, I., Larman, D. G.: Convex bodies, economic cap coverings, random polytopes. *Mathematika* **35**, 274–291 (1988).

[3] Bárány, I., Reitzner, M.: On the variance of random polytopes, *Adv. Math.*, to appear. (2010)

[4] Ewald, G., Larman, D.G., Rogers, C.A.: The directions of the line segments and of the r-dimensional balls on the boundary of a convex body in Euclidean space. *Mathematika* **17**, 1-20 (1970).

[5] Hsu. P. and Robbins, H.: Complete convergence and the law of large numbers. *Proceedings of the National Academy of Sciences, USA, 33*, 25-31 (1947).

[6] Pardon, J.: Central limit theorems for random polygons in an arbitrary convex set. *Annals of Probability (to appear).*

[7] Rényi, A., Sulanke, R.: Über die konvexe Hülle von n zufällig gewählte Punkten. *Z. Wahrsch. Verw. Geb.,* **2**, 75-84 (1963).

[8] Reitzner, M.: Random polytopes and the Efron–Stein jackknife inequality. *Ann. Probab.* **31**, 2136-2166 (2003).

[9] Reitzner, M.: Central limit theorems for random polytopes. *Probab. Theory Relat. Fields* **133**, 483–507 (2005).

[10] Schneider, R., Weil, W.: Stochastic and integral geometry. In: Probability and its Applications. Springer-Verlag, New York-Berlin, 2008.

[11] Schreiber, T., Yukich, J.E.: Variance asymptotics and central limit theorems for generalized growth processes with applications to convex hulls and maximal points, *Ann. Probab.,* **36**, 363–396 (2008).