

Piet Groeneboom

Delft Institute of Applied Mathematics
TU Delft
p.groeneboom@tudelft.nl



Column Piet takes his chance

Convex hulls

Piet Groeneboom regularly writes a column on statistical topics in this magazine.

I recently learned from Geurt Jongbloed that in soccer (British: football) training the convex hull of the team is an important indicator of how well the team is playing. This seems a good excuse to talk about the research on convex hulls of points in the plane. Although with no guarantee that it will be useful for soccer trainers (or rather, with the guarantee that it will not be useful for them).

By the way, there is a ‘Conference on Convex Geometry and Geometric Probability’ in Salzburg, 25–29 September 2023, where Christian Buchta, who himself made important contributions to the field, is one of the main organizers.

In 1988 I proved the following result for the number of vertices N_n of the convex hull of a uniform sample of points from the interior of a convex polygon.

Theorem 1 [7, Corollary 2.5, p. 350]. *Let N_n denote the number of vertices of the boundary of the convex hull of a uniform sample of n points from the interior of a convex polygon with r (≥ 3) vertices. Then*

$$\frac{N_n - \frac{2}{3}r \log n}{\sqrt{\frac{10}{27}r \log n}} \xrightarrow{\mathcal{D}} N(0,1) \quad (1)$$

as $n \rightarrow \infty$, where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution and $N(0,1)$ is the standard normal distribution.

Christian Buchta says in this connection in his paper from 2013 [1]: “A classical result by Rényi and Sulanke [16] states that $\mathbb{E}N_n = \frac{2}{3}r \log n + O(1)$ as $n \rightarrow \infty$. For a quarter of a century, in spite of many efforts, cf., e.g., [9, p. 547] and [10, p. 424], essentially no progress has been achieved concerning the variance of the distribution of N_n .”

This was something I did not know (at the time) that the problem of the variance of N_n was an old unsolved problem. One

would perhaps think that (1) implies that one can say:

$$\text{var}(N_n) \sim \frac{10}{27}r \log n, \quad n \rightarrow \infty,$$

but it could happen that the scaling constant for the central limit theorem is asymptotically different from the square root of the actual variance. I made some frivolous (but — I must say — essentially correct) remarks about this in the introduction of my paper, but was told that mathematicians in the German–Austrian community of stochastic geometers were saying: “Wir glauben es nur wenn wir es sehen.” (or something like that, if I remember correctly from roughly 30 years ago).

You may wonder where the factors $\frac{2}{3}r$ and $\frac{10}{27}r$ come from, not to speak of the $\log n$. There are several answers, depending on the angle from which one approaches the problem. Christian Buchta derives it all in [1] from a (non-trivial) recursive combinatorial relation between expectations for uniform samples of size n and $n + 1$, respectively, where the points lie in the interior of a triangle with vertices $(0,0)$, $(0,1)$ and $(1,0)$.

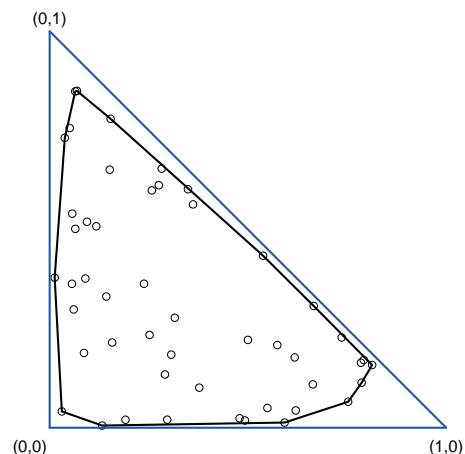


Figure 1 Convex hull of a uniform sample of $n = 50$ points in the interior of a triangle with vertices $(0,0)$, $(0,1)$ and $(1,0)$.

But in this column I mainly want to point attention to the interesting papers of Alexander Nagaev [12] and John Pardon [14,15]. I have the impression that both authors have worked on the problem as outsiders of the field of stochastic geometry (like myself).

I only talked once with Alexander Nagaev (and his brother, I believe) at a conference in Vilnius in 1993, where Alexander explained to me what Isakjan Khamdamov and he had been doing in 1991 in the preprint [13] with my paper from 1988. But for some reason I did (or could) not pay sufficiently close attention to what he was saying then (it was at the conference party).

I shall now first try to describe the ideas of my paper [7]. The number of vertices of the convex hull is the result of a counting process: we walk along the boundary of the convex hull and count the number of vertices we meet in this way. At the time I worked on the problem martingales and the Doob–Meyer decomposition theorem for counting processes and martingales were very ‘hot subjects’, because they could be applied in the theory of right-censored data in statistics. This gave probabilists an argument for stating that their study of martingales, et cetera, was useful in practice and on the other hand statisticians could say that they were applying deep probabilistic tools.

I also ‘went with the wave’ and gave during half a year a seminar on the book *Probabilités et Potentiel* of Paul-André Meyer [11], the original version where Dellacherie was still not co-author) at the Mathematical Centre in Amsterdam in 1977 for a small audience, among whom the statisticians Willem van Zwet, Richard Gill and Kobus Oosterhoff (my own PhD supervisor). It was mainly useful for the topic of the dissertation of Richard Gill [5] (we were sharing a – very small – office at the Mathematical Centre).

So in working on the paper [7], about ten years later, I was curious to see what this theory could do for us in the context of the convex hulls. For example, this result of Rényi and Sulanke for the expectation of the number of vertices, couldn’t we very easily get this from martingale theory?

Let us zoom in on what happens near the corner at (0,0) in Figure 1 and imagine that we can describe the situation there by a Poisson process of points in the first quadrant with intensity 1, for which we compute the left-lower convex hull. For each $a > 0$ we define $W(a) = (U(a), V(a))$ as a point of a realization of the Poisson point process such that there are no points lying strictly to the left of the line $x + ay = c$ through $W(a)$.

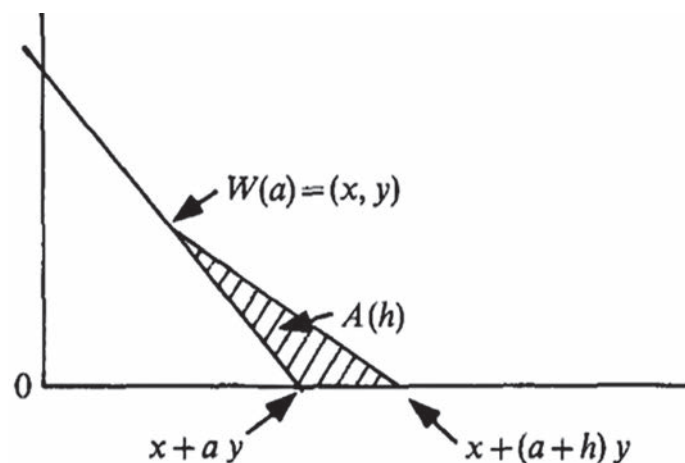


Figure 2 Vertex $W(a)$ of convex hull of Poisson point process in first quadrant.

In fact, we can walk through the left-lower convex hull of our Poisson process by lines $x + ay = c$, $a \in (0, \infty)$, where $c = c(a)$ is chosen so that the line runs through a point $W(a)$ of the point process of this type. This process has no beginning and no ending, but that’s OK. It is even a *Markov process*! That is, if we consider a to be the time parameter, the vertices $W(a')$ we’ll meet for $a' > a$ only depend on $W(a)$ and not on values $W(a')$ with $a' < a$.

We can define $N(a, b)$ as the number of vertices of the convex hull that we meet if we vary the parameter a' of the supporting lines $x + a'y = c(a')$ over the interval $[a, b]$. We then get the following lemma:

Lemma 1 [7, Lemma 2.6, p.342]. *For each $a > 0$, the processes*

$$M_1(a, b) = N(a, b) - \frac{1}{2} \int_a^b V(c)^2 dc, \quad b \geq a, \tag{2}$$

and

$$M_2(a, b) = N(a, b)^2 - \frac{1}{2} \int_a^b \{2N(a, c) + 1\} V(c)^2 dc, \quad b \geq a, \tag{3}$$

are martingales.

In this lemma $V(c)$ is the second coordinate of a vertex $W(c)$ of the convex hull of the Poisson point process.

The way we prove a result like this is illustrated in Figure 2 (a screenshot of [7, Figure 2.4, p.337]). The probability that the number of points in the counting process will be increased by 1 in the time interval $[a, a + h]$ is, as $h \downarrow 0$, equivalent to h times the area of the shaded rectangle in Figure 2, which is $\frac{1}{2}hy^2 = \frac{1}{2}hV(a)^2$. We use here that we are dealing with a Poisson process of intensity 1 in the first quadrant.

This yields, by elementary martingale theory, that the so-called *compensator* of the counting process (also called *predictable projection*) is given by the second term on the right-hand side of (2). Since the expectation of $M_1(a, b)$ does not change if $b \geq a$ increases, the expectation is equal to the expectation at time $b = a$, which is zero. So we get:

$$\mathbb{E}N(a, b) = \frac{1}{2} \int_a^b \mathbb{E}V(c)^2 dc.$$

A trivial one line computation, given in [7, below (2.32)], using the properties of the Poisson process, yields that

$$\frac{1}{2} \mathbb{E}V(c)^2 = \frac{1}{3c},$$

with the conclusion that

$$\mathbb{E}N(a, b) = \frac{1}{3} \log(b/a).$$

In particular, if we choose $a_n = 1/n$ and $b_n = n$, we get:

$$\mathbb{E}N(a_n, b_n) \sim \frac{2}{3} \log n, \quad n \rightarrow \infty.$$

Aha! Here we almost have the result in the Rényi and Sulanke paper! And with so little effort! If we can show that the convex hull of the original sample process near (0,0) behaves in the same way as the convex hull of the Poisson process near (0,0) for the time parameter between $1/n$ and n , and that we have a similar

phenomenon for the other vertices (using affine invariance), and can asymptotically neglect what happens along the sides, we actually have the Rényi and Sulanke result, modulo the transition from Poisson process to sample process. The latter transition is essentially the relation $(1 + \alpha/n)^n \rightarrow e^\alpha$, as John Pardon rightly argues in [15].

Moreover, (3) of Lemma 1 deals with the second moment, and if we use (3), we get information about the variance of N_n , which remained a mystery for a quarter of a century, according to the quote above from Buchta’s paper [1]!

The central limit theorem can be derived from the fact that the process $\{N(a, b), b \geq a\}$ is *strongly mixing*, which can also be deduced from the martingale characterization of the process. This means that the dependence between pieces $N(a, b)$ and $N(a', b')$ if $a < b < a' < b'$ dies out rather rapidly if the distance between b and a' increases.

This program was carried out in [7] and yielded Theorem 1. A similar program was carried out for a uniform sample in the interior of the unit disc. Here we get the following result.

Theorem 2 [7, Theorem 3.4]. *Let N_n denote the number of vertices of the boundary of the convex hull of a uniform sample of n points from the unit disc. Then*

$$\frac{N_n - 2\pi c_1 n^{1/3}}{\sqrt{2\pi c_2 n^{1/3}}} \xrightarrow{\mathcal{D}} N(0, 1)$$

as $n \rightarrow \infty$, $c_1 \approx 0.53846$, $c_2 \approx 0.13160$ [3], and where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution and $N(0, 1)$ is the standard normal distribution.

The constant c_2 has to be evaluated numerically. The constant c_1 had already been evaluated in [16] and can also be written as $c_1 = (\frac{3}{2}\pi)^{-1/3} \Gamma(\frac{5}{3})$. The constant c_2 has been evaluated in [3], using the methods of [7]. It depends on the numerical evaluation of certain multiple integrals and corrects the original computation in [7].

The boundary of the convex hull of a uniform sample in the interior of a circle was studied as a prototype of the situation that the convex set from which the sample is taken has a C^2 boundary with non-vanishing curvature. One sees from Theorem 2 that the order of the number of vertices has increased from $\log n$ to $n^{1/3}$, going from Theorem 1 to Theorem 2. Also note that in Theorem 2 the scaling constants for the expectation and variance of N_n are of the same order, just as in Theorem 1.

The papers of Alexander Nagaev and John Pardon

This whole problem of “are the constants used in the scaling in Theorem 1 above asymptotically the same as the expectation and the square root of the variance of N_n ” was definitely solved by John Pardon in [15]. Corollary 1.2 in [15] gives this result for *all* convex sets K of unit area and at the same time gives a central limit theorem for the area A_n of $K \setminus C_n$, where C_n is the convex hull of a uniform sample from K .

Corollary 1 [15, Corollary 1.2, p.824]. *Let Φ be the normal distribution function. As $n \rightarrow \infty$, the following estimates for the convex hull C_n hold uniformly over all convex sets K of unit area:*

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left\{ \frac{N_n - \mathbb{E}N_n}{\sqrt{\text{var} N_n}} \leq x \right\} - \Phi(x) \right| \rightarrow 0,$$

and

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left\{ \frac{A_n - \mathbb{E}A_n}{\sqrt{\text{var} A_n}} \leq x \right\} - \Phi(x) \right| \rightarrow 0.$$

He derives this from the following fundamental theorem in [14].

Theorem 3 [14, Theorem 1.1]. *Let Φ be the standard normal distribution function. As $\text{area}(K) \rightarrow \infty$, the following estimates hold for the convex hull Π_K of a Poisson point process of intensity 1 in the interior of a convex set K .*

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left\{ \frac{N - \mathbb{E}N}{\sqrt{\text{var} N}} \leq x \right\} - \Phi(x) \right| = O\left(\frac{\log^2 \mathbb{E}N}{\sqrt{\mathbb{E}N}}\right),$$

and

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left\{ \frac{A - \mathbb{E}A}{\sqrt{\text{var} A}} \leq x \right\} - \Phi(x) \right| = O\left(\frac{\log^2 \mathbb{E}A}{\sqrt{\mathbb{E}A}}\right),$$

where N is the number of vertices and A the area of $K \setminus \Pi_K$.

Now I must say that I think that the methods John Pardon uses in [14] are rather close to what I was doing 33 years earlier in [7]. He indeed refers to my paper at several spots in his exposition and also to [2], which gives a central limit theorem for the area. I’ll say more about [2] below.

But [14] has some very nice new twists. First of all there is the parametrization. He does not make the transition to an unbounded Poisson point process, but stays with the Poisson point process inside the original bounded convex set. He also walks through the vertices via points $W(\theta)$, but the parameter θ is now an angle in $\mathbb{R}/2\pi$ and not the parameter a of a line $x + ay = c$. A typical picture in his exposition is given in Figure 3.

The ‘cap’ $C_K(p, \theta)$ in Figure 3(b) is a region above the line with angle θ through $W(\theta)$ which does not contain points of the Poisson point process, parametrized either by a point p on the line segment cut off by the convex set K or parametrized by the area of $C_K(p, \theta)$.

Instead of letting the parameter λ of the Poisson process tend to ∞ and keeping the convex set fixed, it is more natural in his approach to keep the intensity of the Poisson process fixed, for example $\lambda = 1$ (like in the Poisson process considered above) and let the area of the convex set grow to ∞ . This amounts to the same type of asymptotics.

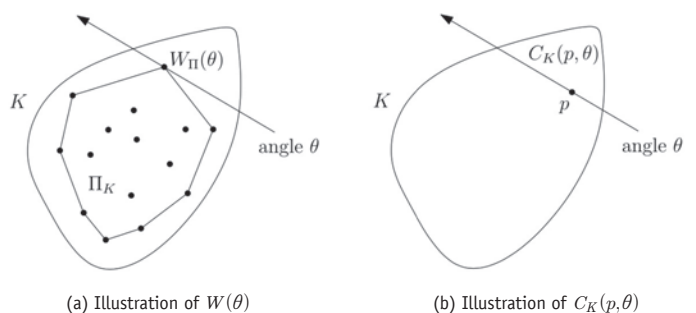


Figure 3 John Pardon’s parametrization.

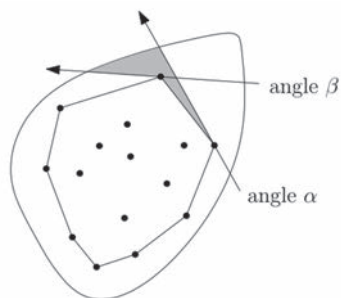


Figure 4 John Pardon's definition of $A(\alpha, \beta)$.

He lets $X(\alpha, \beta)$ be either the number of vertices $N(\alpha, \beta)$ or the area $A(\alpha, \beta)$ one gets by varying the angles of the lines of support between α and β , where $A(\alpha, \beta)$ is the shaded region in Figure 4 (screenshot of [14, Figure 1]).

Then he shows that $X(\alpha, \beta)$ (either number of vertices or area), chopping it up into L pieces $X(\alpha, \alpha_1), X(\alpha_1, \alpha_2), \dots, X(\alpha_{L-1}, \beta)$, corresponds to a strongly mixing process to which he can apply the following result.

Theorem 4 [17]. Let $X = \sum_{i=1}^L X_i$, where X_1, \dots, X_L are random variables. Additionally suppose that

- $\mathbb{E} |X_i|^3 \leq c_1$,
- X_1, \dots, X_L are α -mixing (= strongly mixing) with $\alpha \leq C_2 \exp\{-\delta |i-j|\}$,

for some $\delta > 0$ and $C_1, C_2 < \infty$. Then there exists $M > 0$ such that

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left\{ \frac{X - \mathbb{E}X}{\sqrt{\text{var} X}} \leq x \right\} - \Phi(x) \right| \leq M \frac{L(\log L)^2}{(\text{var} X)^{3/2}}.$$

(Note that the exponent in the power of the denominator of (10.1) in [14] should be $3/2$ instead of $-3/2$.)

This theorem is very interesting! It is from 1991. Had I known this in 1988, this would have spared me the trouble of trying to compute the variance! I derived the right type of strong mixing and also showed that the moment generating function existed in a neighborhood of the origin. Having these results, I could have stated the central limit theorem without bothering about computing the variance!

On the other hand, the variance was this long unsolved problem (something I did not know), and it served as a reference point for Christian Buchta's finite sample result for the number of vertices of the convex hull of a sample from a convex polygon.

John Pardon derives his central limit result from Theorem 4 also by showing the strong mixing property and showing that the moment generating function exists in a neighborhood of the origin. But he does not use martingales, and gets it straight from the properties of the Poisson process.

There are a lot of nice ideas in [14] I could talk about, for example the clever use of compactness arguments and the interesting idea of introducing an affine invariant measure. But the actual purpose of my column is to draw attention to Nagaev's paper [12], which somehow seems to have been neglected.

Alexander Nagaev died tragically in a skiing accident. Tomasz Schreiber was so kind to provide me with the preprint [13]; he is

no longer alive either.

In 2012 I published the paper [6], discussing [13] and [12], and deriving the result of [13] from my own work. But I presently think that my paper [6] did not sufficiently reveal the importance of Nagaev's paper [12].

Unlike Pardon, he works with an unbounded convex set A , like I was doing in the first section and in [7]. He has the following conditions for A :

- a. A is convex unbounded, with $\lambda(A) = \infty$.
- b. The point $(0, 0)$ lies on its boundary.
- c. Each straight line $Z_c = \{(x, y): y = c\}$ cuts off from A a region having finite area for all $c > 0$, i.e., Z_c intersects the boundary of A at two points.

Nagaev considers a doubly infinite sequence of vertices of the convex hull of the Poisson process of intensity 1: $\dots, w_{-2}, w_{-1}, w_0, w_1, w_2, \dots$, with corresponding regions $\dots, \theta_{-2}, \theta_{-1}, \theta_0, \theta_1, \theta_2, \dots$, where θ_0 is the region between the horizontal line through w_0 and the boundary of A , θ_1 the region enclosed by the line through w_0 and w_1 , the horizontal line through w_0 and the boundary of A , θ_{-1} the region enclosed by the line through w_{-1} and w_0 , the horizontal line through w_0 and the boundary of A , et cetera, see Figure 5, a screenshot of Figure 1 in [12].

He then defines the random variables $\xi_j = \lambda(\theta_j)$, the Lebesgue measure of the region θ_j , and γ_- and γ_+ as the left and right intersections of the horizontal line through w_0 with the boundary of A .

Furthermore, he defines for $j \geq 1$ the points γ_j as the intersection of the line through w_{j-1} and w_j with the right boundary of A and for $j \leq -1$ the points γ_j as the intersection of the line through w_j and w_{j+1} with the left boundary of A .

Finally, he defines the length of a line segment with endpoints a and b by $\ell(a, b)$ and then

$$\eta_j = \begin{cases} \ell(\gamma_-, w_0) / \ell(\gamma_-, \gamma_+), & \text{if } j = 0, \\ \ell(w_{j-1}, w_j)^2 / \ell(w_{j-1}, \gamma_j)^2, & \text{if } j > 0, \\ \ell(w_j, w_{j+1})^2 / \ell(\gamma_j, w_{j+1})^2, & \text{if } j < 0, \end{cases}$$

see Figure 5 (a screenshot of [12, Figure 1]). With these definitions he has the following remarkable theorem.

Theorem 5 [13, Theorem 2.1, p.23]. The variables $\xi_j, \eta_k, j, k \in \mathbb{Z}$ are independent. Furthermore, the ξ_j are exponential with param-

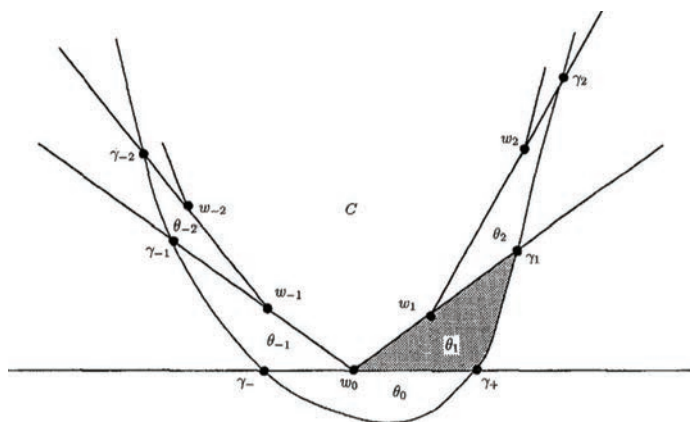


Figure 5 Nagaev's construction.

eter 1 and the η_j are Uniform(0,1).

Note that we can apply this result both for convex regions with C^2 boundaries and for ‘wedges’ like the first quadrant, rotated over the angle $-\pi/4$. For the latter situation we get the following corollary.

Corollary 2 [13, Theorem 1.1]. *Let N_n denote the number of vertices of the boundary of the convex hull of a uniform sample of n points from the interior of a convex polygon K with r (≥ 3) vertices and area equal to 1. Let A_n be the area of $K \setminus C_n$, where C_n is the convex hull of the sample. Then (N_n, A_n) satisfies the following 2-dimensional central limit theorem:*

$$\left(\frac{10}{27}r \log n\right)^{-1/2} \left(N_n - \frac{2}{3}r \log n, nA_n - \frac{2}{3}r \log n\right) \xrightarrow{\mathcal{D}} N(0, \Sigma),$$

where

$$\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & \frac{14}{5} \end{pmatrix}.$$

The proof of the asymptotic normality does not need the strong mixing, which was used by both John Pardon and myself, we just have sums of independent random variables (see, e.g., [6]). On the other hand, the result of John Pardon is still the most general, since it is independent of the type of boundary. Also, in his approach we do not have to go back from the unbounded Poisson point process to the bounded convex figure. Still it seems that some synthesis of these different approaches should perhaps be possible.

I still have to say something on the paper [2]. Here a central limit theorem for the area was derived. Although the method is OK, in the derivation of the constants a time inversion argument was needed that was not applied correctly. The argument with the correct time inversion argument is given in [8]. The matter is also

References

- Christian Buchta, Exact formulae for variances of functionals of convex hulls, *Adv. in Appl. Probab.* 45(4) (2013), 917–924.
- A.J. Cabo and P. Groeneboom, Limit theorems for functionals of convex hulls, *Probab. Theory Related Fields* 100(1) (1994), 31–55.
- Steven Finch and Irene Hueter, Random convex hulls: a variance revisited, *Adv. in Appl. Probab.* 36(4) (2004), 981–986.
- Sh.K. Formanov and I.M. Khamdamov, On joint probability distribution of the number of vertices and area of the convex hulls generated by a Poisson point process, *Statist. Probab. Lett.* 169 (2021), 108966.
- R.D. Gill, *Censoring and Stochastic Integrals*, Mathematical Centre Tracts, Vol. 124, Mathematisch Centrum, Amsterdam, 1980.
- P. Groeneboom, Convex hulls of uniform samples from a convex polygon, *Adv. in Appl. Probab.* 44 (2012), 330–342.
- Piet Groeneboom, Limit theorems for convex hulls, *Probab. Theory Related Fields* 79(3) (1988), 327–368.
- Piet Groeneboom, The remaining area of the convex hull of a Poisson process, 2011, arXiv:1111.2504.
- Nicholas P. Jewell and Joseph P. Romano, Coverage problems and random convex hulls, *J. Appl. Probab.* 19(3) (1982), 546–561.
- Nicholas P. Jewell and Joseph P. Romano, Evaluating inclusion functionals for random convex hulls, *Z. Wahrsch. Verw. Gebiete*, 68(4) (1985), 415–424.
- Paul-André Meyer, *Probabilités et potentiel*, Publications de l’Institut de Mathématique de l’Université de Strasbourg, XIV. Actualités Scientifiques et Industrielles, No. 1318, Hermann, Paris, 1966.
- A.V. Nagaev, Some properties of convex hulls generated by homogeneous Poisson, point processes in an unbounded convex domain, *Ann. Inst. Statist. Math.* 47(1) (1995), 21–29.
- A.V. Nagaev and I.M. Khamdamov, Limit theorems for functionals of random convex hulls, Preprint of Institute of Mathematics, Academy of Sciences of Uzbekistan, Tashkent (in Russian), 1991.
- John Pardon, Central limit theorems for random polygons in an arbitrary convex set, *Ann. Probab.* 39(3) (2011), 881–903.
- John Pardon, Central limit theorems for uniform model random polygons, *J. Theoret. Probab.* 25(3) (2012), 823–833.
- A. Rényi and R. Sulanke, Über die konvexe Hülle von n zufällig gewählten Punkten, *Z. Wahrsch. Verw. Gebiete* 2 (1963), 75–84.
- Ī Sunklodas, Approximation of distributions of sums of weakly dependent random variables by the normal distribution, in *Probability Theory* 6 (Russian), Itogi Nauki i Tekhniki, Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., 1991, pp. 140–199.