

Pacific Journal of Mathematics

ON THE DISTRIBUTION OF PYTHAGOREAN TRIANGLES

JOACHIM LAMBEK AND LEO MOSER

ON THE DISTRIBUTION OF PYTHAGOREAN TRIANGLES

J. LAMBEK AND L. MOSER

1. Introduction. This paper was conceived with the object of estimating the number $P_a(n)$ of primitive Pythagorean triangles with *area* less than n . The problem seemed of interest, since F. L. Miksa [5] recently tabulated all primitive Pythagorean triangles with area less than 10^9 , in order of increasing area. The method employed here also yields known estimates for the numbers $P_h(n)$ and $P_p(n)$ of primitive Pythagorean triangles with *hypotenuse* and *perimeter*, respectively, less than n ; we use $P(n)$ as generic notation for all of these.

D. N. Lehmer [4] had shown in 1900 that

$$P_h(n) \sim \frac{1}{2} \pi^{-1} n, \quad P_p(n) \sim \log 2 \cdot \pi^{-2} n.$$

In 1948, D. H. Lehmer [3] obtained

$$P_p(n) = \log 2 \cdot \pi^{-2} n + O(n^{1/2} \log n),$$

pointing out that this disproved a conjecture of Krishnaswami [2] that $P_p(n) \sim n/7$. For primitive Pythagorean triangles with area less than $2 \cdot 10^6$, W. P. Whitlock [6] found that

$$\left| P_a(n) - \frac{1}{2} n^{1/2} + 5 \right| \leq 2.$$

However, Miksa's table, which goes 500 times as far as Whitlock's, suggested that $P_a(n)$ is not asymptotic to $(1/2)n^{1/2}$.

In § 2 we reduce the problem of approximating $P(n)$ to that of estimating the number of lattice points in certain regions of the Cartesian plane. The latter problem is treated in § 3, with some attempt at generality. In § 4 we obtain the following asymptotic formulae for $P(n)$:

$$P_h(n) = \frac{1}{2} \pi^{-1} n + O(n^{1/2} \log n),$$

Received June 16, 1953. This paper was written at the Summer Research Institute of the Canadian Mathematical Congress.

Pacific J. Math., 5 (1955), 73-83

$$P_p(n) = \log 2 \cdot \pi^{-2} n + O(n^{1/2} \log n),$$

$$P_a(n) = cn^{1/2} + O(n^{1/3}),$$

where $c = \Gamma(1/4)^2 2^{-1/2} \pi^{-5/2} = .531340\dots$.

Let $E(n) = cn^{1/2} - P_a(n)$. The following table, constructed on the basis of Miksa's tabulations, gives an idea of the possible constant suggested by $E(n) = O(n^{1/3})$:

$10^{-8} n$	$P_a(n)$	$cn^{1/2}$	$E(n)$	$E(n)n^{-1/3}$
1	5157	5313.4	138	.298
2	7342	7514.3	172	.294
3	9007	9203.1	196	.295
4	10405	10627	222	.301
5	11644	11908	237	.299
6	12778	13015	237	.282
7	13800	14058	258	.291
8	14755	15029	274	.296
9	15655	15940	285	.295
10	16513	16802	289	.289

For the foregoing data, the average value of $E(n)n^{-1/3}$ is .295 with a standard deviation of .00405. We are led to conjecture that $E(n) \sim c'n^{1/3}$, where c' is approximately .295.

2. Pythagorean triangles. A Pythagorean triangle

$$\Delta = \langle a, b, c \rangle = \langle b, a, c \rangle$$

is determined by three positive integers a, b, c such that $a^2 + b^2 = c^2$. If their greatest common divisor $(a, b, c) = 1$, then Δ is called *primitive*. If $(a, b, c) \leq 2$, we shall call Δ *quasi-primitive*. An integral lattice point $\langle x, y \rangle$ on the Cartesian plane will be called *primitive* if $(x, y) = 1$.

LEMMA 1. *The equations*

$$(1) \quad a = 2xy, \quad b = x^2 - y^2, \quad c = x^2 + y^2$$

determine a one-to-one correspondence between all quasi-primitive Pythagorean

triangles $\Delta = \langle a, b, c \rangle$ and all primitive lattice points $\langle x, y \rangle$ of the region $x > y > 0$ of the Cartesian plane.

Proof. If a, b, c are given by (1), then clearly Δ is a Pythagorean triangle. Moreover, if $(x, y) = 1$, Δ will be quasi-primitive. It is known [1, p. 189] that all primitive Pythagorean triangles are uniquely expressible in the form (1) with $(x, y) = 1$ and $x \not\equiv y \pmod{2}$. It remains to consider the case $(a, b, c) = 2$. Then $\langle a/2, b/2, c/2 \rangle$ is primitive, and we may write

$$a/2 = x'^2 - y'^2, \quad b/2 = 2x'y', \quad c/2 = x'^2 + y'^2,$$

where

$$x' > y' > 0, \quad (x', y') = 1, \quad x' \not\equiv y' \pmod{2}.$$

If we now let $x = x' + y'$, $y = x' - y'$ and eliminate x', y' in favor of x, y , we may easily verify (1) and $x > y > 0$, $(x, y) = 1$. This completes the proof.

In the following, let $F(\Delta) = F(a, b, c)$ be homogeneous of degree $k > 0$ in a, b, c , such that there are only finitely many Δ with $F(\Delta) < n$. Without loss in generality we may assume that F has been normalized so that

$$(2) \quad F(\Delta) \geq 1.$$

In this paper, we are interested in three special cases:

Case 1. $F(\Delta) = c$ (hypotenuse), $k = 1$.

Case 2. $F(\Delta) = a + b + c$ (perimeter), $k = 1$.

Case 3. $F(\Delta) = ab/2$ (area), $k = 2$.

It is seen that Condition (2) is satisfied in these cases.

We wish to find the number $P(n)$ of primitive and the number $Q(n)$ of quasi-primitive Pythagorean triangles Δ for which $F(\Delta) < n$. Now $F(2a, 2b, 2c) < n$ if and only if $F(a, b, c) < n/2^k$; hence

$$Q(n) = P(n) + P(n/2^k).$$

This formula may be inverted to give

$$(3) \quad P(n) = \sum_{i \geq 0} (-1)^i Q(n/2^{ki}).$$

It should be noted that this is a finite sum, since, by (2), $Q(1) = 0$.

The calculation of $P(n)$ is thus reduced to that of $Q(n)$. By Lemma 1, $Q(n)$ is the number of primitive lattice points in the region of the Cartesian plane defined by the inequalities

$$(4) \quad G(x, y) = F(2xy, x^2 - y^2, x^2 + y^2) < n, \quad x > y > 0.$$

If we write $n = t^{2k}$, this is the same as the region

$$G(x/t, y/t) < 1, \quad x > y > 0,$$

that is, the set of all points $\langle x, y \rangle$ for which $\langle x/t, y/t \rangle$ lies in the region R defined by

$$(5) \quad G(X, Y) < 1, \quad X > Y > 0.$$

If R is any subset of the Cartesian plane, and t any positive real number, we define Rt to be the region obtained from R by radial magnification in the ratio $t:1$, so that $\langle x, y \rangle$ lies in Rt if and only if $\langle x/t, y/t \rangle$ lies in R . Furthermore, let $L(R)$ denote the number of integral lattice points in R , $L'(R)$ the number of primitive lattice points in R .

In particular, if R is the region defined by (5), it follows that (4) is the region Rt , so that

$$(6) \quad Q(n) = L'(Rt), \quad n = t^{2k}.$$

For any R , $\langle x, y \rangle$ is an integral lattice point in Rt with $(x, y) = i$ if and only if $\langle x/i, y/i \rangle$ is a primitive lattice point in Rt/i . Hence

$$(7) \quad L(Rt) = \sum_{i \geq 1} L'(Rt/i).$$

To avoid questions of convergence, we shall confine our attention to the case

$$(8) \quad L(R) = 0, \quad L'(R) = 0,$$

these two conditions being clearly equivalent. In particular, if R is the region defined by (5), then (8) follows from (2). The expression on the right of (7) is now a finite sum, and may be inverted with the help of the well-known Möbius function $\mu(i)$ [1, p. 236] to give

$$(9) \quad L'(Rt) = \sum_{i \geq 1} \mu(i) L(Rt/i).$$

This again is a finite sum.

In view of (3), (6), and (9), the problem of calculating $P(n)$ has been reduced to that of counting the number of lattice points in the region Rt , where \bar{R} is given by (5).

3. On the number of lattice points in a region. Let R be an open set in the Cartesian plane. We wish to approximate the number $L(Rt)$ of integral lattice points in Rt by the measure

$$M(Rt) = M(R)t^2$$

of Rt . Here, as before, Rt denotes the region obtained from R by radial magnification in the ratio $t:1$.

Instead of fixing the lattice and magnifying R in the ratio $t:1$, we may keep R fixed and shrink the mesh of the lattice in the ratio $1:t$. Let L_t denote the lattice thus contracted, with mesh length $1/t$. Then $L(Rt)$ is also the number of vertices of L_t in R .

LEMMA 2. *If R is the open region enclosed by a simple closed Jordan curve in the Cartesian plane, whose total horizontal plus vertical variation is V , then*

$$|L(Rt) - M(Rt)| \leq Vt.$$

Proof. Let L_t^* be the lattice conjugate to L_t , that is, the square lattice whose vertices are the midpoints of the squares of L_t . Then each vertex of L_t in R lies in a square of L_t^* which has a part in common with R , and each square of L_t^* contained in R contains a vertex of L_t . Let $s_t(R)$ denote the number of closed squares of L_t^* contained in R , and $S_t(R)$ the number of open squares of L_t^* having a part in common with R ; then

$$s_t(R) \leq L(Rt) \leq S_t(R).$$

Moreover, comparing areas, we obtain

$$s_t(R)t^{-2} \leq M(R) \leq S_t(R)t^{-2},$$

so that

$$|L(Rt) - M(R)t^2| \leq S_t(R) - s_t(R).$$

Now this is the number of open squares of L_t^* which contain portions of the given Jordan curve J , hence does not exceed the number of horizontal and vertical lines of L_t^* crossed by J . As there are t mesh lengths of L_t^* per unit interval, the latter number is bounded by Vt , where V is the total horizontal plus vertical variation of J . This completes the proof.

We wish to obtain a result analogous to Lemma 2 for unbounded regions. It seems difficult to state the most general result of this kind. Here we confine our attention to the following:

LEMMA 3. *Let R be the region in the Cartesian plane defined in polar coordinates ρ, ϑ by the inequalities*

$$0 < \rho < f(\vartheta), \quad 0 \leq \alpha < \vartheta < \beta \leq \pi/2,$$

subject to:

- (i) $f(\vartheta)$ is continuous, increasing, and positive for $\alpha \leq \vartheta < \beta$,
- (ii) $f(\vartheta) \asymp (\beta - \vartheta)^{\mu-1}$, $1 > \mu > 1/2$,
- (iii) $\tan \beta$ is rational.

Then

$$L(Rt) - M(Rt) = O(t^{1/\mu}).$$

Proof. The distance from a point on the curve $\rho = f(\vartheta)$ to the line $\vartheta = \beta$ is given by

$$(10) \quad g(\vartheta) = f(\vartheta) \sin(\beta - \vartheta) \asymp (\beta - \vartheta)^\mu,$$

which tends to 0 as $\vartheta \rightarrow \beta$, since $\mu \geq 0$. On the other hand,

$$f(\vartheta) \asymp (\beta - \vartheta)^{\mu-1}$$

tends to infinity, since $\mu < 1$. Hence the line $\vartheta = \beta$ is an asymptote of the curve.

We shall write

$$\tan \beta = p/q, \quad (p, q) = 1, \quad p^2 + q^2 = r^2.$$

The distance from a point $\langle x, y \rangle$ below the asymptote to the asymptote is then

* The symbol of $f \asymp g$ is used to denote $0 < \underline{\lim} f/g \leq \overline{\lim} f/g < \infty$.

$$\rho \sin (\beta - \vartheta) = x \sin \beta - y \cos \beta = (px - qy)/r.$$

Hence the smallest nonzero distance which any integral lattice point can have from the asymptote is $1/r$, and the smallest nonzero distance from a vertex of L_t to the asymptote is $1/(rt)$.

For sufficiently large t , we have

$$g(\alpha) > \frac{1}{2rt}.$$

Since $g(\vartheta) \rightarrow 0$, we can choose a ϑ_t such that

$$(11) \quad g(\vartheta_t) = f(\vartheta_t) \sin(\beta - \vartheta_t) = \frac{1}{2rt},$$

and $g(\vartheta) < 1/(2rt)$ for $\vartheta > \vartheta_t$. Let R_t be the region defined by

$$0 < \rho < f(\vartheta), \quad \vartheta_t \leq \vartheta < \beta;$$

then R_t contains no vertices of L_t ; that is, $L(R_t t) = 0$. Hence

$$L(R_t) - M(R_t) = L(R_t - R_t t) - M(R_t - R_t t) - M(R_t t),$$

so that

$$(12) \quad |L(R_t) - M(R_t)| \leq V_t t + M(R_t t),$$

by Lemma 2, if V_t denotes the total (horizontal plus vertical) variation of the boundary of $R - R_t$. It remains to estimate $V_t t$ and $M(R_t t)$.

We claim that $V_t = O(f(\vartheta_t))$. For the boundary of $R - R_t$ consists of two straight segments of lengths $f(\alpha)$ and $f(\vartheta_t)$ and the arc $\rho = f(\vartheta)$, $\alpha < \vartheta < \vartheta_t$. We need only consider the variation of the latter. Its vertical variation is the variation of $f(\vartheta) \sin \vartheta$. Now this is an increasing function of ϑ , and hence has variation

$$O(f(\vartheta_t) \sin \vartheta_t) = O(f(\vartheta_t)).$$

The horizontal variation of the arc is the variation of $f(\vartheta) \cos \vartheta$. Now this can be expressed as the difference of two increasing functions $f(\vartheta)$ and $f(\vartheta)(1 - \cos \vartheta)$, both of whose variations are $O(f(\vartheta_t))$. Hence so is the horizontal variation and therefore also the total variation of the arc, as was to be proved.

From (10) and (11) we obtain

$$(13) \quad (\beta - \vartheta_t)^\mu \asymp t^{-1};$$

hence

$$V_t t = O(f(\vartheta_t) (\beta - \vartheta_t)^{-\mu}) = O(\beta - \vartheta_t)^{-1} = O(t^{1/\mu}),$$

as required. Finally,

$$M(R_t t) = M(R_t) t^2 = \frac{1}{2} t^2 \int_{\vartheta_t}^{\beta} f(\vartheta)^2 d\vartheta = O\left(t^2 \int_{\vartheta_t}^{\beta} (\beta - \vartheta)^{2\mu-2} d\vartheta\right),$$

by (ii). Since $\mu > 1/2$, this is

$$O((\beta - \vartheta_t)^{2\mu-1} t^2) = O((\beta - \vartheta_t)^{-1}) = O(t^{1/\mu}),$$

by (13). In view of (12), this completes the proof of Lemma 3.

4. Distribution of Pythagorean triangles. We shall obtain asymptotic formulae for $Q(n)$ and $P(n)$ in the three cases under consideration.

Case 1. Estimation of $P_h(n)$. Here $F(a, b, c) = c$, $k = 1$; and R is given by

$$x^2 + y^2 < 1, \quad x > y > 0.$$

Clearly $M(R) = \pi/8$. Lemma 2 yields

$$L(Rt) = M(R) t^2 + O(t).$$

Hence, by (9),

$$(14) \quad \begin{aligned} L'(Rt) &= \sum_{1 \leq i \leq t} \mu(i) L(Rt/i) \\ &= \sum_{1 \leq i \leq t} \{ \mu(i) M(R) (t/i)^2 + O(t/i) \} \\ &= M(R) t^2 \{ 6\pi^{-2} + O(t^{-1}) \} + O(t \log t) \\ &= \frac{3}{4} \pi^{-1} t^2 + O(t \log t). \end{aligned}$$

Then (6) becomes

$$Q_h(n) = \frac{3}{4} \pi^{-1} n + O(n^{1/2} \log n),$$

and (3) gives rise to

$$\begin{aligned} P_h(n) &= \sum_{i \geq 0} (-1)^i Q(n/2^i) \\ &= \frac{3}{4} \pi^{-1} n \sum_{i \geq 0} (-1/2)^i + O\left(\sum_{i \geq 0} (n/2^i)^{1/2} \log(n/2^i)\right) \\ &= \frac{1}{2} \pi^{-1} n + O(n^{1/2} \log n), \end{aligned}$$

as stated in § 1.

Case 2. Estimation of $P_p(n)$. Here $F(a, b, c) = a + b + c$, $k = 1$, and R is given by

$$2x(x + y) < 1, \quad x > y > 0.$$

By integration, $M(R) = (\log 2)/4$. Calculating as in Case 1, we obtain

$$Q_p(n) = \frac{3}{2} \log 2 \cdot \pi^{-2} n + O(n^{1/2} \log n),$$

and

$$P_p(n) = \log 2 \cdot \pi^{-2} n + O(n^{1/2} \log n),$$

as stated in § 1.

Case 3. Estimation of $P_d(n)$. Here $F(a, b, c) = ab/2$, $k = 2$; and R is given by

$$xy(x^2 - y^2) < 1, \quad x > y > 0.$$

Transformed into polar coordinates, this becomes

$$\rho^4 \sin 4 \vartheta < 4, \quad 0 < \vartheta < \pi/4.$$

By integration,

$$M(R) = 2^{-5/2} \pi^{-1/2} \Gamma(1/4)^2.$$

The line $\vartheta = \pi/8$ separates R into two subregions R_1 and R_2 , which we shall take to be open sets, R_1 with asymptote $\vartheta = 0$, and R_2 with asymptote $\vartheta = \pi/4$. R_2 satisfies the conditions of Lemma 3, with $\mu = 3/4$. Although Lemma 3 does not apply directly to R_1 , it may be used for the reflection of R_1 about the line $\vartheta = \pi/4$. Such a reflection does not affect the area of R_1 or the number of lattice points in it. Again we have $\mu = 3/4$. Hence

$$L(R_i t) = M(R_i t) + O(t^{4/3}) \quad (i = 1, 2).$$

Adding these two equations, and observing that there are no vertices of L_t on the dividing line $\vartheta = \pi/8$, its slope being irrational, we obtain

$$L(Rt) = M(R)t^2 + O(t^{4/3}).$$

Hence, by (9),

$$\begin{aligned} L^*(Rt) &= \sum_{i \geq 1} \mu(i) L(Rt/i) \\ &= \sum_{i \geq 1} \{ \mu(i) M(R) (t/i)^2 + O(t/i)^{4/3} \} \\ &= 6\pi^{-2} M(R) t^2 + O(t^{4/3}). \end{aligned}$$

Then (6) becomes

$$Q_a(n) = 6\pi^{-2} M(R) n^{1/2} + O(n^{1/3}),$$

so that, by (3),

$$\begin{aligned} (15) \quad P_a(n) &= \sum_{i \geq 0} (-1)^i Q(n/4^i) \\ &= 6\pi^{-2} M(R) n^{1/2} \sum_{i \geq 0} (-1/2)^i + O\left(\sum_{i \geq 0} (n/4^i)^{1/3} \right) \\ &= 4\pi^{-2} M(R) n^{1/2} + O(n^{1/3}). \end{aligned}$$

Replacing $M(R)$ by its numerical value, we obtain the result stated in § 1.

REFERENCES

1. G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, Oxford, 1945.
2. A. A. Krishnaswami, *On isoperimetrical Pythagorean triangles*, Tôhoku Math. J. **27** (1926), 332 - 348.
3. D. H. Lehmer, *A conjecture of Krishnaswami*, Bull. Amer. Math. Soc. **54** (1948), 1185 - 1190.
4. D. N. Lehmer, *Asymptotic evaluation of certain totient sums*, Amer. J. Math. **22** (1900), 293 - 335.
5. F. L. Miksa, *Table of primitive Pythagorean triangles with area up to 10^9 in order of increasing area*, unpublished.
6. W. P. Whitlock Jr., *Rational right triangles with equal area*, Scripta Math. **9** (1943), 155 - 167 and 265 - 268.

MCGILL UNIVERSITY,
UNIVERSITY OF ALBERTA

Frank Herbert Brownell, III, <i>Flows and noncommuting projections on Hilbert space</i>	1
H. E. Chrestenson, <i>A class of generalized Walsh functions</i>	17
Jean Bronfenbrenner Crockett and Herman Chernoff, <i>Gradient methods of maximization</i>	33
Nathan Jacob Fine, <i>On groups of orthonormal functions. I</i>	51
Nathan Jacob Fine, <i>On groups of orthonormal functions. II</i>	61
Frederick William Gehring, <i>A note on a paper by L. C. Young</i>	67
Joachim Lambek and Leo Moser, <i>On the distribution of Pythagorean triangles</i>	73
Roy Edwin Wild, <i>On the number of primitive Pythagorean triangles with area less than n</i>	85
R. Sherman Lehman, <i>Approximation of improper integrals by sums over multiples of irrational numbers</i>	93
Emma Lehmer, <i>On the number of solutions of $u^k + D \equiv w^2 \pmod{p}$</i>	103
Robert Delmer Stalley, <i>A modified Schnirelmann density</i>	119
Richard Allan Moore, <i>The behavior of solutions of a linear differential equation of second order</i>	125
William M. Whyburn, <i>A nonlinear boundary value problem for second order differential systems</i>	147