Bertrand 'paradox' reloaded (with details on transformations of variables)*

G. D'Agostini
Università "La Sapienza" and INFN, Roma, Italia
(giulio.dagostini@roma1.infn.it, http://www.roma1.infn.it/~dagos)

Abstract

This note is just to point out, if needed, that uncertainty about models and their parameters has little to do with a 'paradox'. The proposed 'solution' is to formulate practical questions instead of seeking refuge into abstract principles. (And, in order to be concrete, some details on how to calculate the probability density functions of the chord lengths are provided, together with some comments on simulations.)

" On trace au hasard une corde dans un cercle.
Quelle est la probabilité pour qu'elle soit plus petite
que le côté du triangle équilatéral inscrit?
. . .
Entre ces trois réponses, quelle est la véritable?
Aucune des trois n'est fausse, aucune n'est exacte,
la question est mal posée."
(Joseph Bertrand)

" Probability is either referred to real cases
or it is nothing"
(Bruno de Finetti)

"As far as the laws of mathematics refer to reality,
they are not certain,
and as far as they are certain,
they do not refer to reality."
(Albert Einstein)

*Note based on lectures to PhD students in Rome. More on the subject, including the Android app mentioned in the text, at http://www.roma1.infn.it/~dagos/Bertrand.html.

1 Introduction

The question asked by Joseph Bertrand in his 1889 book *Calcul des probabilités* [1] is about the probability that a chord drawn 'at random' is smaller that the side of the equilateral triangle inscribed in the same circle (see e.g. [2]). Obviously the question can be restated asking about the probability that the chord will be larger or smaller than the radius, or whatever segment you like, upper to the diameter (for which the solution is trivially 100%). The reason of the original choice of the side of such a triangle is that the calculation is particularly easy, *under the hypotheses* Bertrand originally considered, as we shall.

The question can be restated in more general terms, i.e. that of finding the probability distribution of length l of a chord. Indeed, as well known, our uncertainty about the value a continuous variable can assume can be described by a *probability density function*, hereafter 'pdf', $f(l)$, which should be written, more precisely, as $f(l\,|\,I_s(t))$, where I is the *Information* available to the *subject* s at the *time* t. In fact,

> "*Since the knowledge may be different with different persons or with the same person at different times, they may anticipate the same event with more or less confidence, and thus different numerical probabilities may be attached to the same event.*" [3]

And, hence, probability is always conditional probability, as again well stated by Schrödinger [3],

> "*Thus whenever we speak loosely of 'the probability of an event,' it is always to be understood: probability with regard to a certain given state of knowledge.*"

These quotes, which are 100% in tune with common sense, definitely rule out to use the appellative of 'paradox' for the problem of the chords. In other words, Bertrand's 'paradox' belongs to a completely different class than e.g. Bertrand Russell's *barber paradox*. Absurd is instead the positions of those who maintain that the problem should have a unique solution once it is *"well posed"*'[4] or that they have found the *"conclusive answer"* [5].

In my point of view the question proposed by Bertrand can be only answered if framed in a given contest and 'asked' somehow, either to human beings, or – and hence the quote marks – to Nature by performing suitable experiments (but making a particular simulation, of the kind of that proposed in Ref. [5], is the same as asking human beings – and even making an experiment the result will depend on the set up!).

For example we can ask suddenly students, without any apparent reason (for them), to draw a chord in a circle printed on a sheet of paper. And to give more sense (and fun) to the 'experiment' we can make a bet among us on the resulting length, in units of the radius of the circle (the bet could be even more detailed, concerning for example the orientation of the chord – it never happened to me that a student drawn a vertical one, but perhaps Japanese students might have higher tendency to draw segments top down!). Or we can ask students to write, with the their preferred programming language and plotting package, a 'random chord generator'. In this case our bet will be about the length that will result from a certain extraction, e.g. the first, or the 100th (if we were not informed about all or some of the previous 99 results, because this information might change our odds about the outcome of the following ones).

Indeed I have done experiments of these kinds since many years, and I have formed my opinion on how students react, depending on the class they attend and how they are skilled in mathematics, and . . . even in games (yes! in the answer there is even a flavor of game theory, because smart students unavoidable try to guess the reason of the question and try to surprise you!) For example unsophisticated students draw 'typical' chords, of the kind you can get searching for the keywords "`chord geometry`" in Google Images.[1] Hence, for example, a real 'well posed' question could be the following: "what is the length of the chord (in units of the radius) that will appear in e.g. the 27-th (from left to right, top to bottom) image returned by search engine?" When instead I propose the the question to students of advanced years I have quite some expectation that one or more of them will draw a diameter (just a maximum chord) or even a tiny one almost tangent to the circumference.

Essentially this is all what I have to say about this so called 'paradox'. The rest of the note has been written for didactic purposes, in order to show how to evaluate the probability distributions of interest and how to make the simulations.

2 Four 'basic' solutions

Let us now start going through the 'classical' solutions of the problem, i.e. those analyzed by Bertrand and which typically appear in the literature and on the web, plus a forth one which is even (somehow) simpler than the other three.

[1] `https://www.google.it/search?q=chord+geometry&source=lnms&tbm=isch`

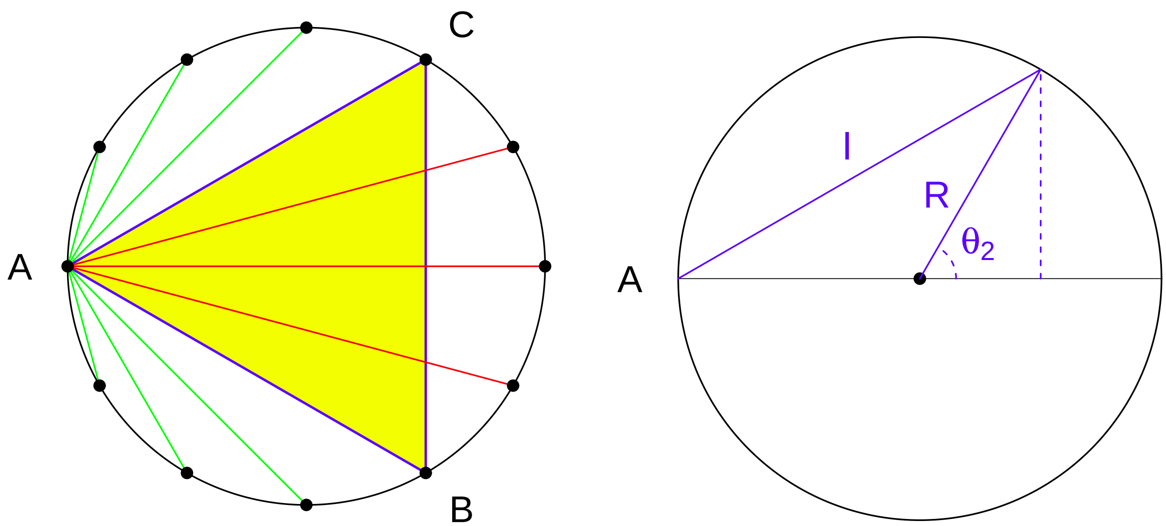

Figure 1: Left: circle with inscribed equilateral triangle and chords with one end in A and the other distributed uniformly around the circumference. Right: geometric construction to show how to evaluate the length l from θ_2, with $0 \leq \theta_2 \pi$ (see text for details).

2.1 Endpoints uniformly chosen on the circumference

A way to draw 'at random' a chord is to choose two points on the circumference and to join them with a segment. If we indicate the first point with A, corresponding to a vertex of the equilateral triangle, as shown in Fig. 1, the chords smaller that the side of the triangle are those with the other end is either in the arc between A and B or in that between A and C. The resulting probability is thus simply **2/3**.

A more complete information about our beliefs that the length falls in any given interval is provided by the pdf $f(l \,|\, \mathcal{M}_1)$, where $\mathcal{M}_1$ stands for 'Model 1'. Since there is a correspondence between a point on the circle and the angle between the radius to that point and the x-axis (according to usual trigonometry convention), we can turn our condition into two angles, θ_1 and θ_2, uniformly distributed between 0 and 2π. If we are only interested in the length of the chords and not in their position inside the circle we can fix θ_1 at π and consider θ_2 in the interval between 0 and π. The corresponding chord will have a length of (see right plot in Fig. 1)

$$\begin{aligned} l &= \sqrt{(R + R\cos\theta_2)^2 + R^2\sin^2\theta_2} \\ &= R\sqrt{(1+\cos\theta_2)^2 + \sin^2\theta_2} \\ &= R\sqrt{2+2\cos\theta_2}\,, \end{aligned} \qquad (1)$$

or, more conveniently, the normalized length λ will be

$$\lambda = \frac{l}{R} \;=\; \sqrt{2+2\cos\theta_2}\,. \tag{2}$$

The problem is thus how to calculate the pdf[2] $f(\lambda\,|\,\mathcal{M}_1)$ from

$$f(\theta_2\,|\,\mathcal{M}_1) \;=\; \frac{1}{\pi} \qquad (0 \le \theta_2 \le \pi)\,. \tag{3}$$

We shall use the general rule[3]

$$f(y) \;=\; \int_{-\infty}^{+\infty} \delta\,(y-g(x)) \cdot f(x)\,dx\,, \tag{4}$$

[2]Obviously, the pdf's $f(\theta_2\,|\,\mathcal{M}_1)$ and $f(\lambda\,|\,\mathcal{M}_1)$ are usually expressed by different mathematical functions, a point very clear among physicists. Mathematics oriented guys like to clarify it, thus writing e.g. $f_\Lambda(\lambda\,|\,\mathcal{M}_1)$, $f_{\Theta_2}(\theta_2\,|\,\mathcal{M}_1)$, and so on. I will add a proper subscript only if it is not clear from the context what is what.

[3]An alternative way is to use the 'text book' transformation rule, valid for a monotonic function $y=g(x)$ that relates the generic variable X to the variable Y:

$$f_Y(y) \;=\; f_X(g^{-1}(y)) \left|\frac{d}{dy} g^{-1}(y)\right|\,, \qquad (A)$$

which can be derived in the following way for the general variables X and Y [capital letters indicates the variable, small letters the possible values – now it becomes important to make clear the different pdf's and we shall then use the notation $f_X()$ and $f_Y()$]:

$\boxed{g'(x) \ge 0}$ If $g()$ is *non-decreasing* in the range of X we have

$$F_Y(y) \equiv P(Y \le y) = P(g(X) \le y) = P(X \le g^{-1}(y)) \equiv F_X(g^{-1}(y))\,.$$

Making use of the rules of calculus we have then

$$f_Y(y) = \frac{d}{dy}\,F_Y(y) = \frac{d}{dy}\,F_X(g^{-1}(y)) = f_X((g^{-1}(y))\cdot\frac{d}{dy}\,g^{-1}(y)$$

$\boxed{g'(x) \le 0}$ If, instead, $g()$ is *non-increasing* we have

$$F_Y(y) \equiv P(Y \le y) = P(g(X) \le y) = P(X \ge g^{-1}(y)) \equiv 1 - F_X(g^{-1}(y))\,,$$

and then

$$f_Y(y) = \frac{d}{dy}\,F_Y(y) = -\frac{d}{dy}\,F_X(g^{-1}(y)) = -f_X((g^{-1}(y))\cdot\frac{d}{dy}\,g^{-1}(y) = f_X((g^{-1}(y))\cdot\left[-\frac{d}{dy}\,g^{-1}(y)\right]$$

where the factorization in the last step is due to the fact that $f_X()$ cannot be negative, and for this reason the absolute value in (A) is only on the second factor.

Equation (A) takes then into account the two possibilities and, let us stress once more, it is valid for monotonic transformations. We shall use it, to double check our results in footnotes 5, 6 and 7.

in which Y is related to the X by $Y = g(X)$, with $g()$ a generic function and $\delta()$ the Dirac delta. In our case we have then

$$f(\lambda \,|\, \mathcal{M}_1) \;=\; \int_0^{\pi} \delta\left(\lambda - \sqrt{2+2\cos\theta_2}\right) \cdot f(\theta_2 \,|\, \mathcal{M}_1)\, d\theta_2 \,, \tag{5}$$

Eq. (5) has the very simple interpretation of 'summing up' all elements '$f(\theta_2 \,|\, \mathcal{M}_1)\, d\theta_2$' that contribute to the 'same value' of λ (the quote marks are due to the fact that we are dealing with continuous quantities and hence we have use the rules of calculus). This interpretation is very useful to estimate $f(\lambda \,|\, \mathcal{M}_1)$ by simulation: extract 'many' values of θ_2 according to $f(\theta_2 \,|\, \mathcal{M}_1)$; for each value of θ_2 calculate the corresponding λ; summarize the result by suitable statistical indicators and visualize it with an histogram.

Making use of the properties of the Dirac delta and taking into account that $\lambda(\theta_2)$ decreases monotonically we get [4]

$$f(\lambda \,|\, \mathcal{M}_1) \;=\; \int_0^{\pi} \delta\left(\lambda - \sqrt{2+2\cos\theta_2}\right) \cdot \frac{1}{\pi}\, d\theta_2 \tag{6}$$

[4] In the step from Eq. (6) to Eq. (7) we are making use of the famous property (at least among physicists) of the Dirac delta

$$\delta(g(x)) \;=\; \sum_i \frac{\delta(x-x_i)}{|g'(x_i)|}\,,$$

where x_i are the real roots of $g(x)$. In our case we have a single root, which we write as x^*, and hence we get

$$\delta(g(x)) \;=\; \frac{\delta(x-x^*)}{|g'(x^*)|}\,.$$

If we apply it to the general transformation rule (4) we obtain

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} \delta(y-g(x)) \cdot f_X(x)\, dx \\ &= \int_{-\infty}^{\infty} \frac{\delta(x-x^*)}{|g'(x^*)|} \cdot f_X(x)\, dx \\ &= \frac{1}{|g'(x^*)|} \cdot f_X(x^*)\,, \end{aligned}$$

in which we recognize Eq. (A) of footnote 3, if we note that $f_X(x^*)$ is $f_X(g^{-1}(y))$ and the derivative of $g(x)$ w.r.t. x calculated in x^* is the inverse of the derivative of the inverse function g^{-1} (w.r.t. y!) calculated in $y = g(x^*)$ [for the latter observation just think at the Leibniz notation $dy/dx = 1/(dx/dy)$].

Anyway, in order to avoid confusion the denominator in Eq. (7) has been written in the most unambiguous way.

$$= \int_0^\pi \frac{\delta(\theta_2 - \theta_2^*)}{\left| \left(\frac{d}{d\theta_2}(\lambda - \sqrt{2 + 2\cos\theta_2}) \right) \Big|_{\theta_2=\theta_2^*} \right|} \cdot \frac{1}{\pi}\, d\theta_2\,, \tag{7}$$

where θ_2^* is the solution of the equation

$$\lambda - \sqrt{2 + 2\cos\theta_2} \;=\; 0\,, \tag{8}$$

that is

$$\theta_2^* \;=\; \arccos\left(\frac{\lambda^2}{2} - 1\right)\,. \tag{9}$$

Being the derivative in Eq. (7)

$$\left(\frac{d}{d\theta_2}(\lambda - \sqrt{2 + 2\cos\theta_2})\right)\Bigg|_{\theta_2=\theta_2^*} \;=\; \frac{\sin\theta_2}{\sqrt{2 + 2\cos\theta_2}}\Bigg|_{\theta_2=\theta_2^*} \tag{10}$$

$$= 1 - \left(\frac{\lambda}{2}\right)^2 \tag{11}$$

the pdf of interest is [5]

$$f(\lambda\,|\,\mathcal{M}_1) \;=\; \frac{1}{\pi} \cdot \frac{1}{\sqrt{1 - (\lambda/2)^2}} \qquad (0 \le \lambda \le 2)\,, \tag{12}$$

from which we can also calculate the cumulative function

$$F(\lambda\,|\,\mathcal{M}_1) \equiv \int_0^\lambda f(\lambda'\,|\,\mathcal{M}_1)\, d\lambda' \;=\; \frac{2}{\pi} \cdot \arcsin\left(\frac{\lambda}{2}\right)\,, \tag{13}$$

both shown in Fig. 2.

[5] As an exercise, we can check the result with that obtainable using the 'text book' transformation rule described in footnote 3. In our case, using the general symbol $g()$ introduced there, we have $\lambda = g(\theta_2) = \sqrt{2 + 2\cos\theta_2}$ and $\theta_2 = g^{-1}(\lambda) = \arccos\left(\lambda^2/2 - 1\right)$. Applying Eq. (A) of footnote 3 we have

$$\begin{aligned} f_\Lambda(\lambda) &= f_{\Theta_2}(g^{-1}(\lambda)) \cdot \left| \frac{d}{d\lambda} g^{-1}(\lambda) \right| \\ &= \frac{1}{\pi} \cdot \left| \frac{d}{d\lambda} \arccos\left(\lambda^2/2 - 1\right)) \right| \\ &= \frac{1}{\pi} \cdot \left| -\frac{\lambda}{\sqrt{1 - (\lambda^2/2 - 1)^2}} \right|\,, \end{aligned}$$

also yielding Eq. (12). (The minus sign resulting from the derivation is because λ decreases as θ_2 increases, as clear from Fig 1.)

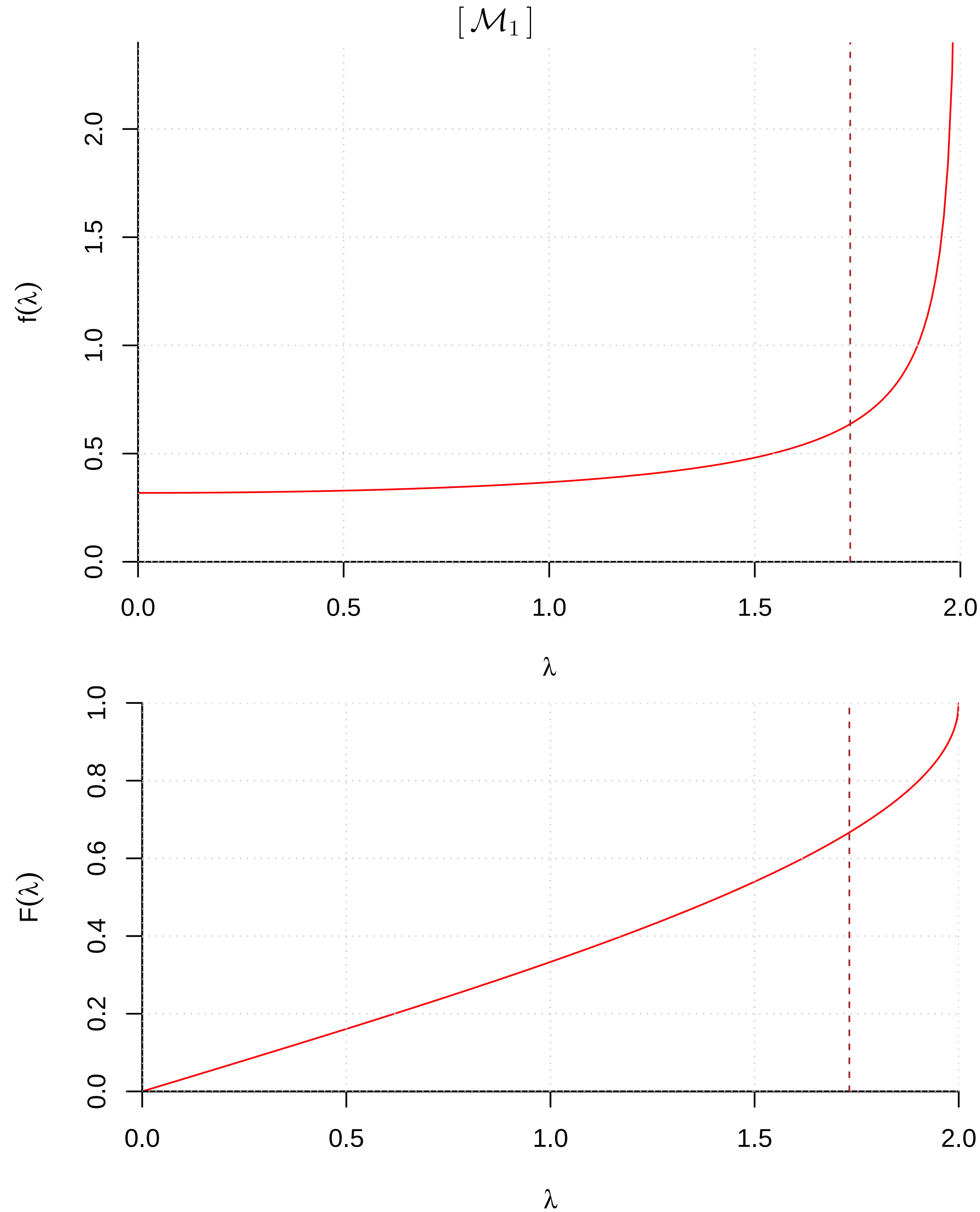

Figure 2: Probability distribution of $\lambda = l/R$ of the chords generated with Method 1. The dashed vertical line indicates $\lambda = \sqrt{3}$.

We can finally check the probability of interest, and also calculate the probability of a chord to be smaller than the radius of the circle:

$$\begin{aligned} P(\lambda \leq \sqrt{3}\,|\,\mathcal{M}_1) = F(\sqrt{3}\,|\,\mathcal{M}_1) &= \frac{2}{\pi}\cdot \arcsin\left(\frac{\sqrt{3}}{2}\right) = \frac{2}{\pi}\cdot\frac{\pi}{3} = \mathbf{\frac{2}{3}} \\ P(\lambda \leq 1\,|\,\mathcal{M}_1) = F(1\,|\,\mathcal{M}_1) &= \frac{2}{\pi}\cdot \arcsin\left(\frac{1}{2}\right) = \frac{2}{\pi}\cdot\frac{\pi}{6} = \frac{1}{3} \end{aligned}$$

Simulations of chords with this methods are reported in the top left plot of Figures 4-7. Figure 4 shows a sample of random chords. Figure 6 shows the position of the center of the chords and, finally, Fig. 5 is the histogram of the normalized lengths. Figure 7 plot shows, as an extra curiosity, the distribution of the distance of the chords from the center of the circle.

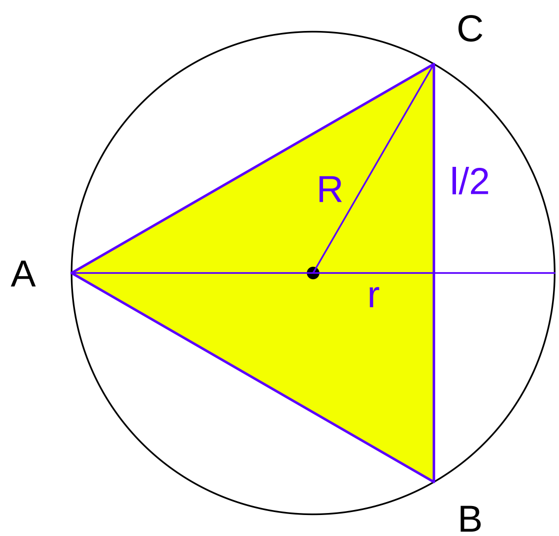

Figure 3: Construction of a chord orthogonal to a radius and distant r from the center of the circle.

2.2 Chords orthogonal to a radius, with center uniformly distributed along it

The second ‘classical’ algorithm consists in choosing chords orthogonal to a radius with its center uniformly distributed along it. As we easily see from Fig. 3, the condition for a chord to be smaller than the side of the triangle (l_T) is that its

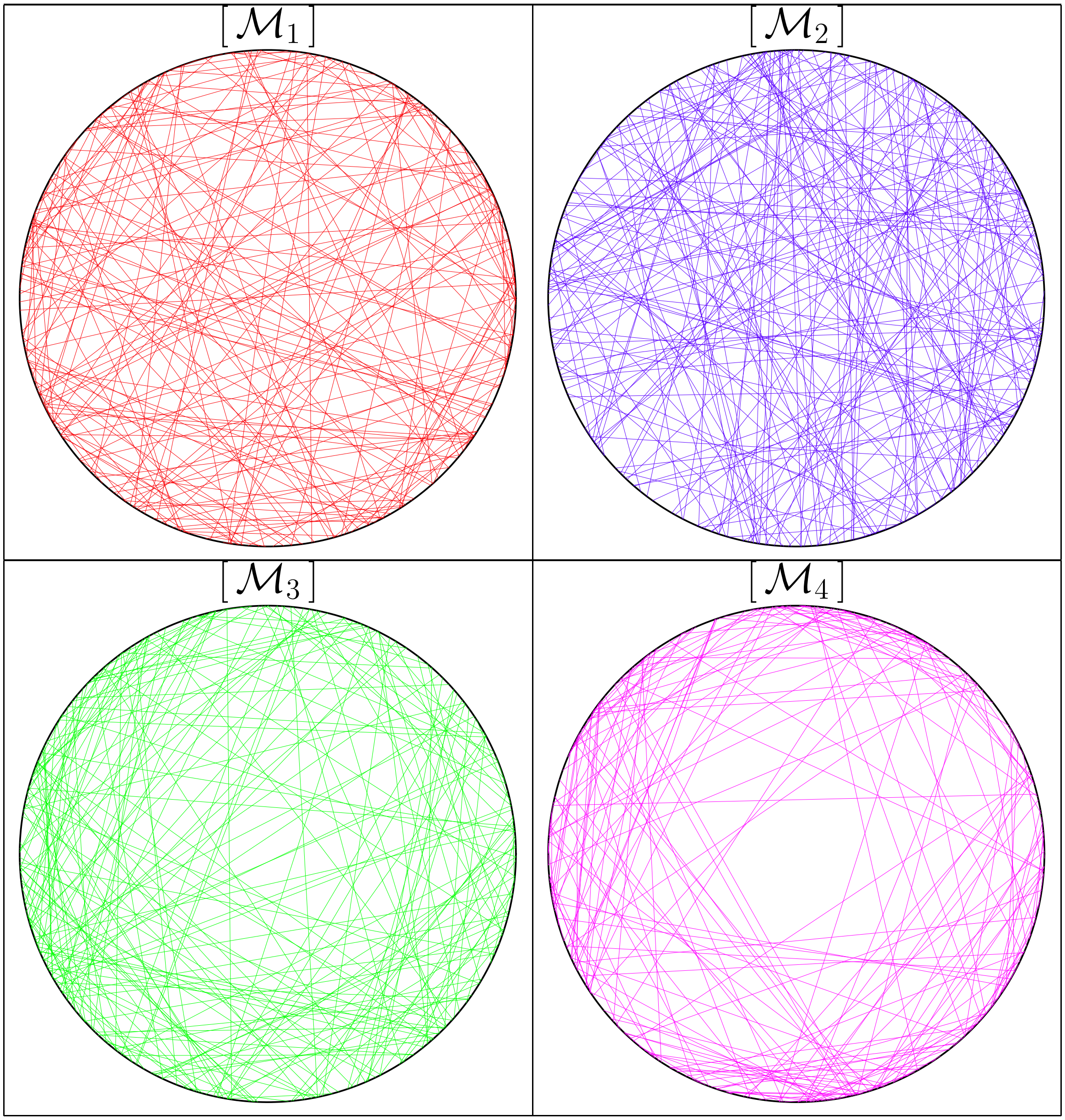

Figure 4: Samples of chords generated the 4 'basic' methods.

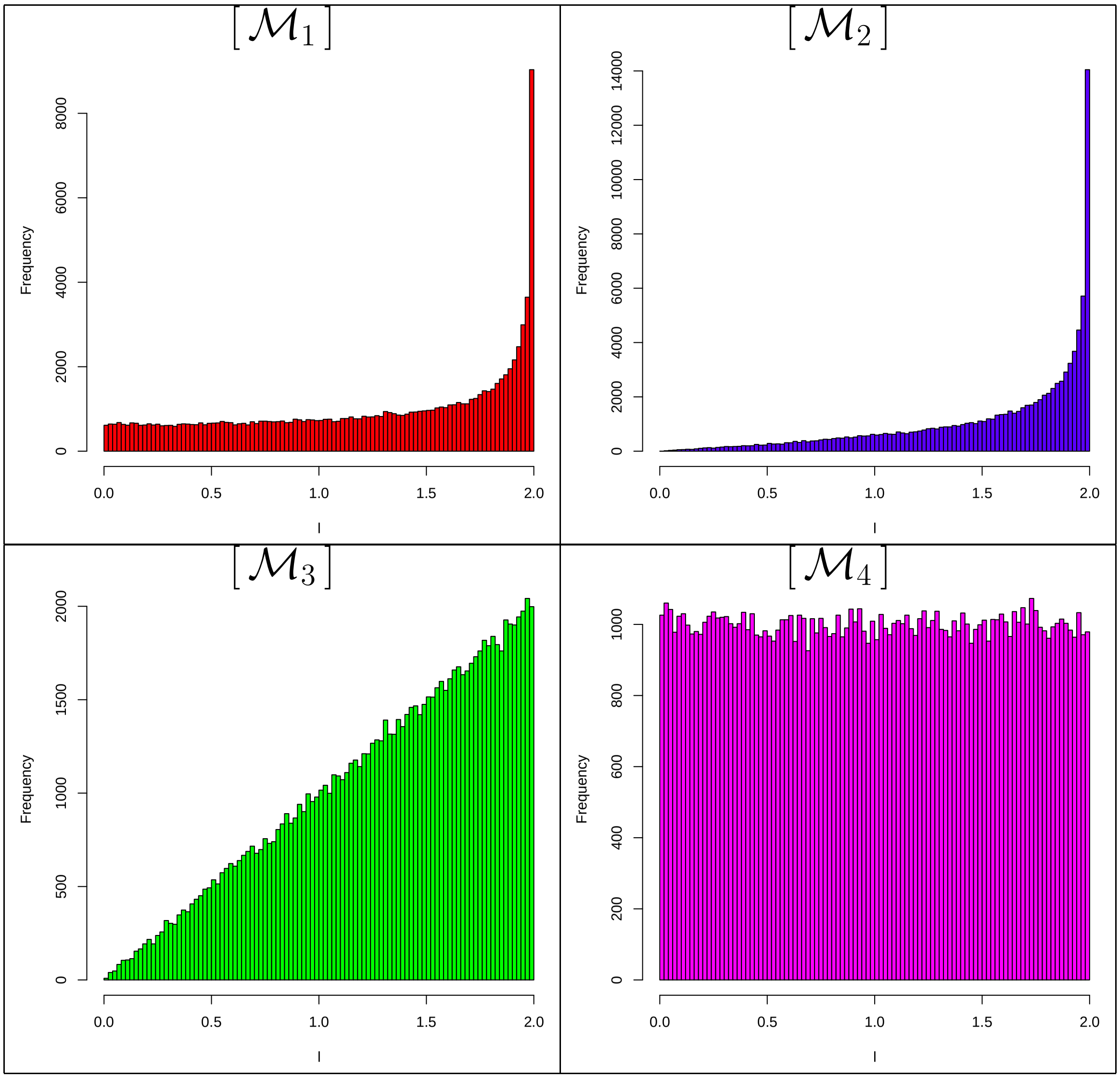

Figure 5: Distribution of the length of the chords in sample produced with the four 'basic' methods.

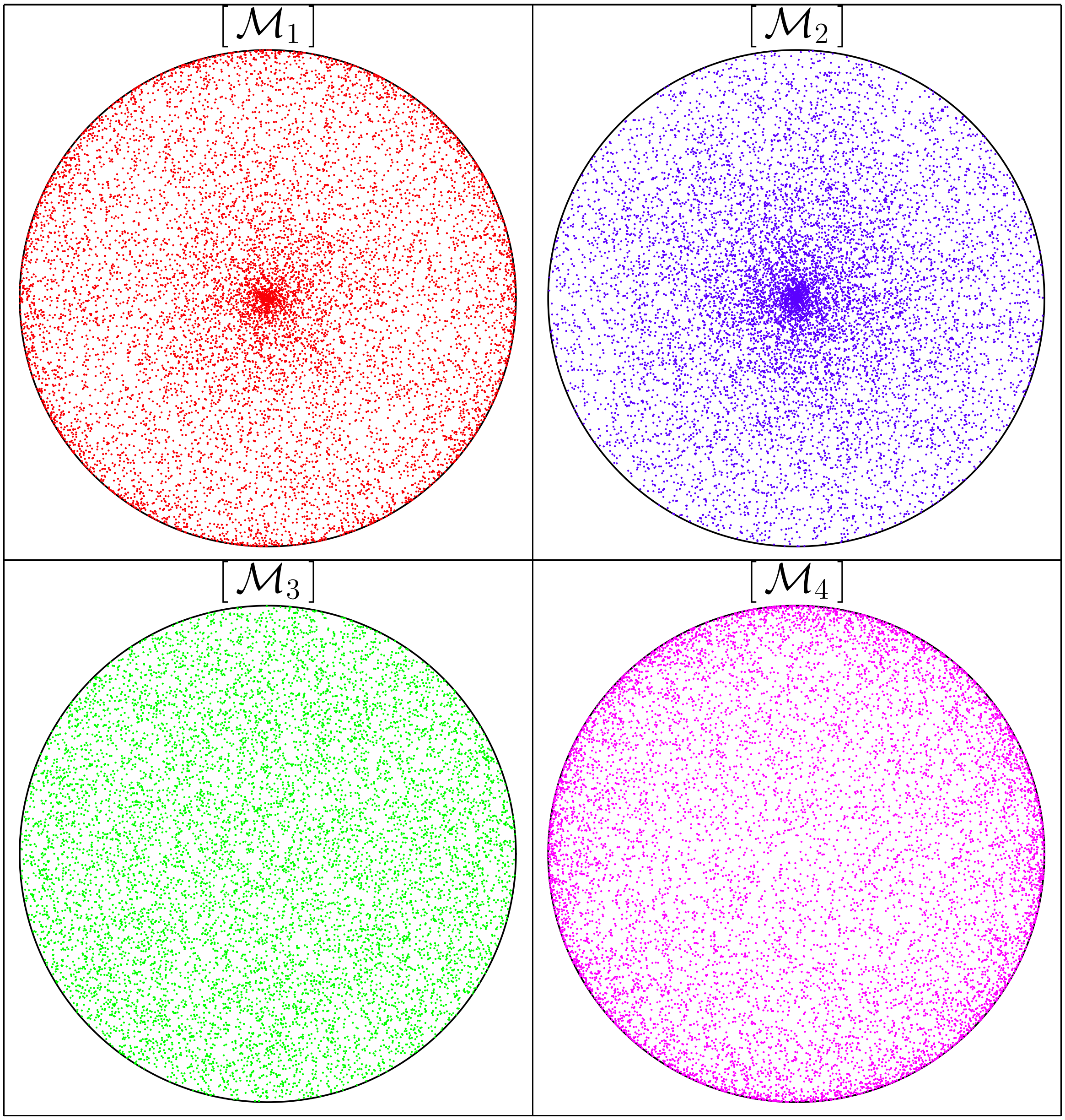

Figure 6: Sample of centers of the chords generated with the four 'basic' methods

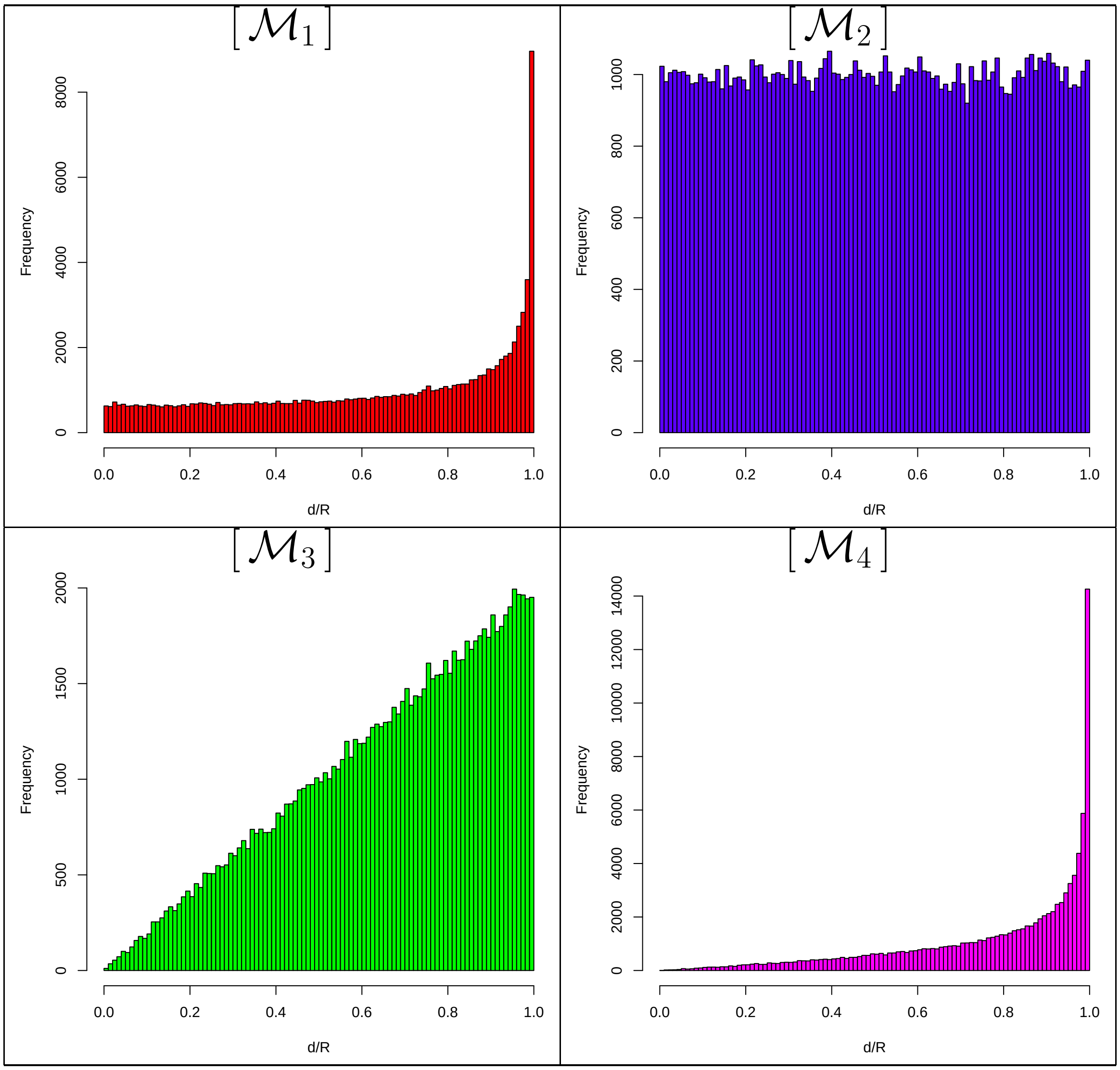

Figure 7: Distribution of the distance of the chords from center of in sample produced with the four 'basic' methods.

distance from the center of the circle, indicated by r in the figure, is above $R/2$. That is

$$P(l \leq l_T \,|\, \mathcal{M}_2) \;=\; P(r \geq R/2 \,|\, \mathcal{M}_2) = \mathbf{\frac{1}{2}}\,. \tag{14}$$

Let us repeat the exercise of evaluating the probability distribution of the lengths of the chords obtained with this method. Using ρ to indicate r/R, in analogy to $\lambda = l/R$, we have (see figure)

$$\lambda \;=\; 2\sqrt{1-\rho^2}\,, \tag{15}$$

with

$$f(\rho \,|\, \mathcal{M}_2) \;=\; 1 \qquad (0 \leq \rho \leq 1)\,. \tag{16}$$

Then, the pdf of interest will be given by

$$f(\lambda \,|\, \mathcal{M}_2) \;=\; \int_0^1 \delta\left(\lambda - 2\sqrt{1-\rho^2}\right)\cdot 1\, d\rho \tag{17}$$

$$=\; \int_0^1 \frac{\delta(\rho-\rho^*)}{\left|\left(\frac{d}{d\rho}(\lambda - 2\sqrt{1-\rho^2})\right)\Big|_{\rho=\rho^*}\right|}\, d\rho\,, \tag{18}$$

$$=\; \frac{1}{2\rho^*/\sqrt{1-\rho^{*2}}} \tag{19}$$

with

$$\rho^* \;=\; \sqrt{1-(\lambda/2)^2}\,. \tag{20}$$

The pdf and the cumulative distribution of interest are then [6]

$$f(\lambda \,|\, \mathcal{M}_2) \;=\; \frac{\lambda}{4\sqrt{1-(\lambda/2)^2}} \qquad (0 \leq \lambda \leq 2) \tag{21}$$

$$F(\lambda \,|\, \mathcal{M}_2) \;=\; 1 - \sqrt{1-(\lambda/2)^2}\,, \tag{22}$$

[6] Let us repeat the exercise of using Eq. (A) of footnote 3 also in this case, starting now from $\lambda = 2\sqrt{1-\rho^2} \equiv g(\rho)$, $\rho = \sqrt{1-(\lambda/2)^2} \equiv g^{-1}(\lambda)$ and $f_P(\rho) = 1$:

$$f_\Lambda(\lambda) \;=\; f_P(g^{-1}(\lambda)) \cdot \left|\frac{d}{d\lambda}\sqrt{1-(\lambda/2)^2}\right|$$

$$=\; 1 \cdot \frac{\lambda}{4\sqrt{1-(\lambda/2)^2}}\,,$$

that is precisely Eq. (21).

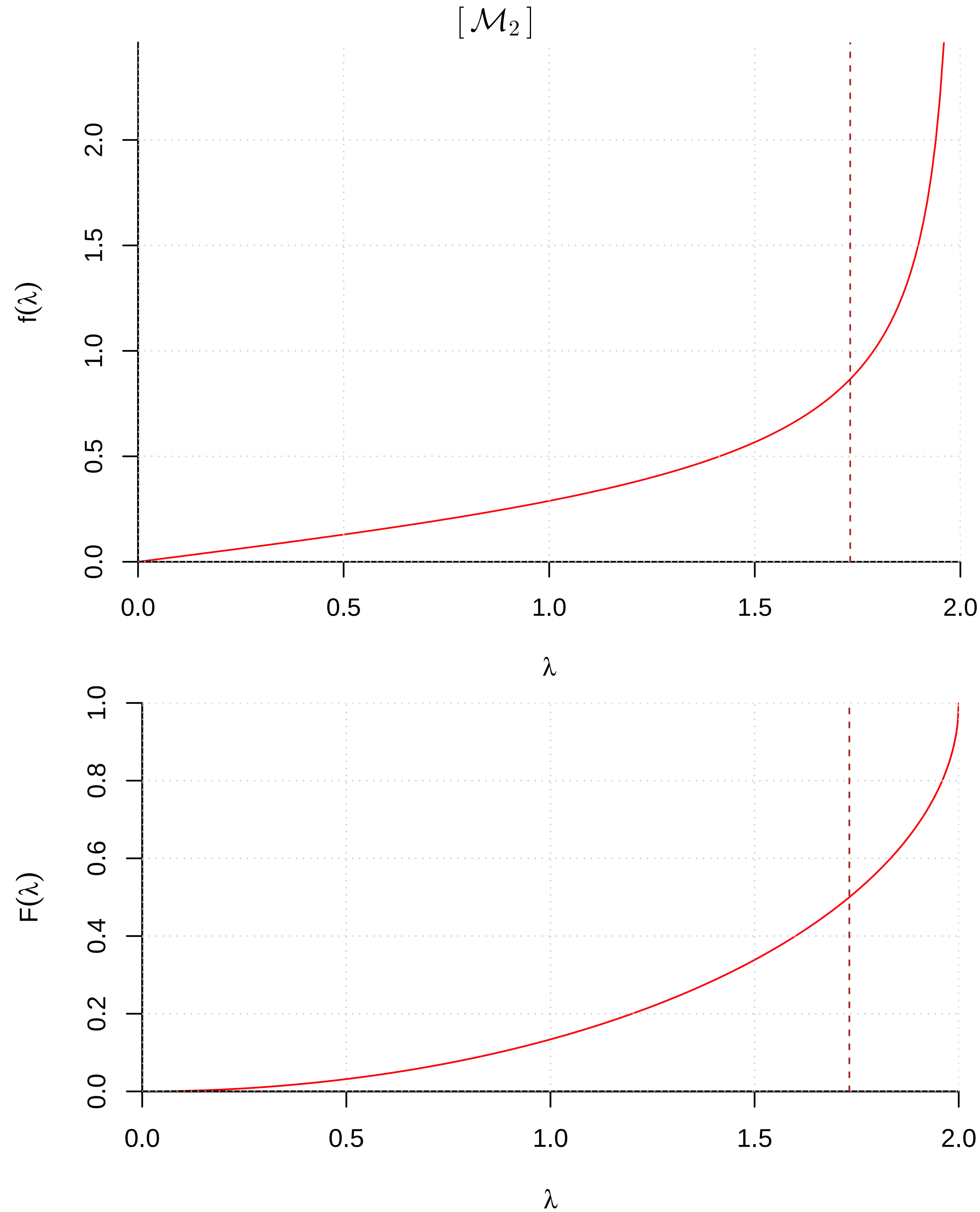

Figure 8: Probability distribution of $\lambda = l/R$ of the chords generated with Method 2. The dashed vertical line indicates $\lambda = \sqrt{3}$.

plotted in Fig. 8 and from which we can calculate the probabilities of interest:

$$F(\sqrt{3}\,|\,\mathcal{M}_2) \;=\; \mathbf{\frac{1}{2}} \tag{23}$$

$$F(1\,|\,\mathcal{M}_2) \;=\; 1-\sqrt{3}/2 \approx 13.4\%. \tag{24}$$

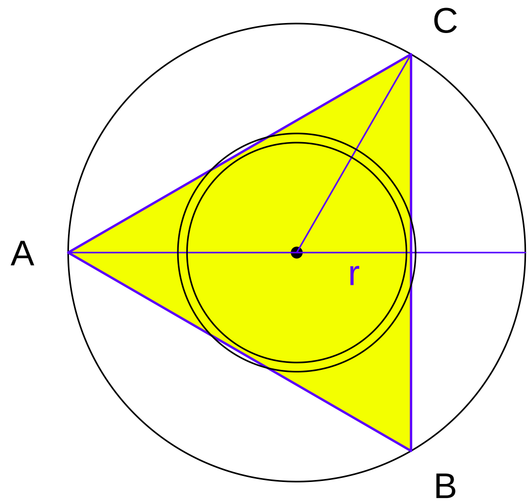

Figure 9: As Fig. 3, but with the annulus of infinitesimal width dr drawn at r to show that in Model 3 the infinitesimal probability dP that the distance of a chord from the center is between r and $r+dr$ is given by $dP \equiv dF(r) = 2\pi r dr/(\pi R^2) = (2r/R^2)\,dr$, and hence $dF(\rho) = 2\rho\,d\rho$, or, in our notation, $f(\rho\,|\,\mathcal{M}_3) = 2\rho$, with $\rho = r/R$.

2.3 Center of chords uniformly chosen inside the circle, with chords orthogonal to radius

The third method is a variant of the second, in which the center of the chord, instead of being generated uniformly along a radius, are generated uniformly inside the circle. The centers of the chords fall inside the circle of radius $R/2$ with probability $1/4$, and than (see again Fig. 3)

$$P(l \le l_T\,|\,\mathcal{M}_3) \;=\; P(r \ge \frac{R}{2}\,|\,\mathcal{M}_3) = \mathbf{\frac{3}{4}}\,. \tag{25}$$

Let us repeat once more the exercise of calculating the probability distribution of λ. The difference with respect to the previous case is that now the pdf of the center of the chords is proportional to r, since $dP \propto (2\pi r)dr$ (see Fig. 9). The pdf of ρ is then, after proper normalization,

$$f(\rho\,|\,\mathcal{M}_3) \;=\; 2\rho \qquad (0 \le \rho \le 1)\,. \tag{26}$$

and the equivalent of Eqs. (17)-(18) and sequel are now

$$
\begin{aligned}
f(\lambda \,|\, \mathcal{M}_3) &= \int_0^1 \delta\left(\lambda - 2\sqrt{1-\rho^2}\right)\cdot 2\rho\, d\rho && (27)\\
&= \int_0^1 \frac{\delta(\rho-\rho^*)\cdot 2\rho}{\left|\left(\frac{d}{d\rho}(\lambda - 2\sqrt{1-\rho^2})\right)\Big|_{\rho=\rho^*}\right|}\, d\rho\,, && (28)\\
&= \frac{2\,\rho^*}{2\rho^*/\sqrt{1-\rho^{*2}}} = \sqrt{1-\rho^{*2}} && (29)
\end{aligned}
$$

with the same ρ^* of Eq. (20), thus leading to [7]

$$
f(\lambda \,|\, \mathcal{M}_3) = \sqrt{1-(1-(\lambda/2)^2)} = \frac{\lambda}{2}\,. \qquad (30)
$$

The result can be then summarized as

$$
\begin{aligned}
f(\lambda \,|\, \mathcal{M}_3) &= \frac{\lambda}{2} \qquad (0 \le \lambda \le 2) && (31)\\
F(\lambda \,|\, \mathcal{M}_3) &= \frac{\lambda^2}{4}\,, && (32)
\end{aligned}
$$

from which we obtain

$$
\begin{aligned}
F(\sqrt{3} \,|\, \mathcal{M}_3) &= \mathbf{\frac{3}{4}} && (33)\\
F(1 \,|\, \mathcal{M}_3) &= \frac{1}{4}. && (34)
\end{aligned}
$$

[7] Let us repeat once more the exercise done in footnotes 5 and 6, since in this case the starting pdf is not a constant, being $f_P(\rho) = 2\rho$. All the rest is like in footnote 6. Here it is:

$$
\begin{aligned}
f_\Lambda(\lambda) &= f_P(g^{-1}(\lambda))\cdot\left|\frac{d}{d\lambda}\sqrt{1-(\lambda/2)^2}\right|\\
&= 2\sqrt{1-(\lambda/2)^2}\cdot\frac{\lambda}{4\sqrt{1-(\lambda/2)^2}} = \frac{\lambda}{2}\,.
\end{aligned}
$$

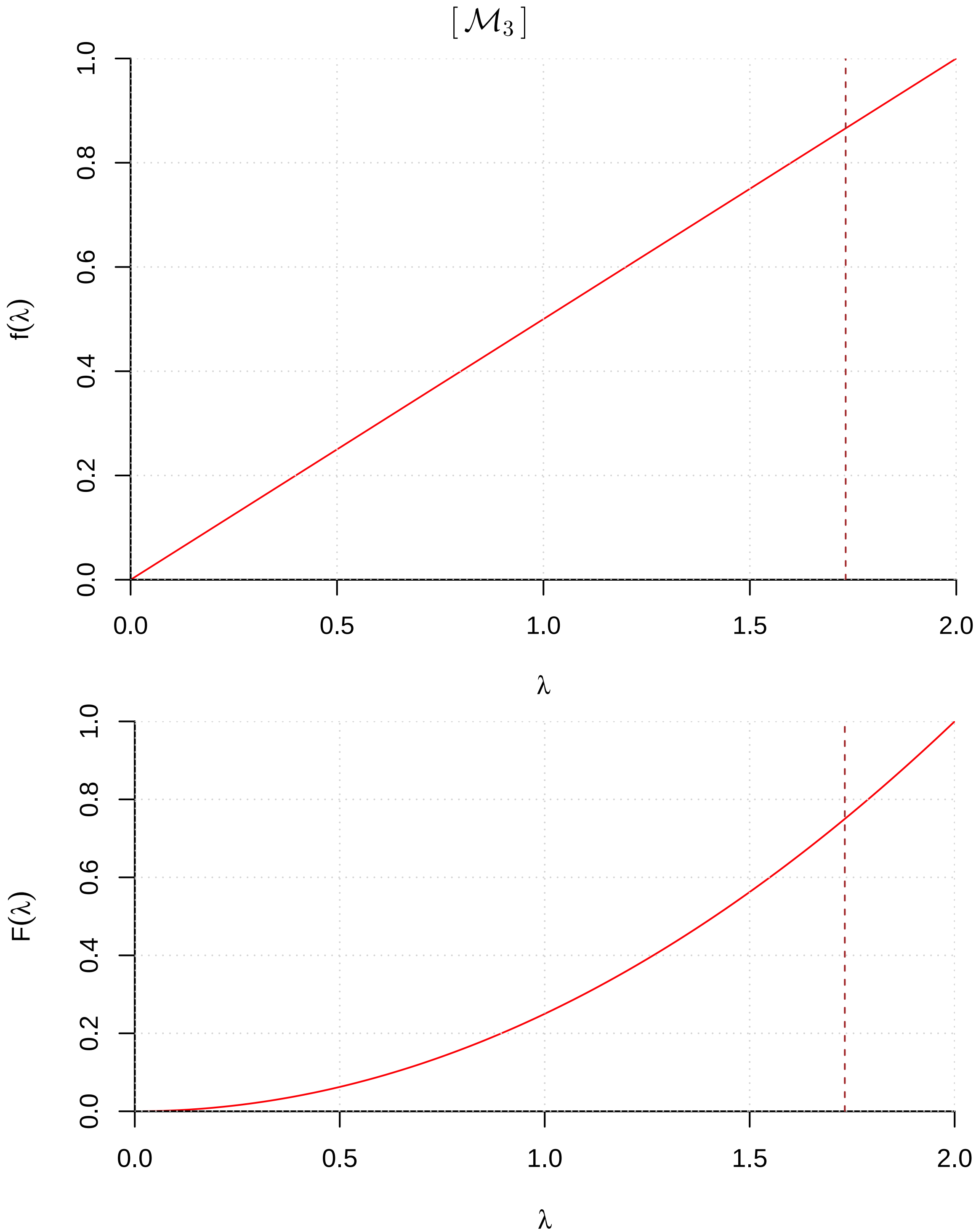

Figure 10: Probability distribution of $\lambda = l/R$ of the chords generated with Method 3. The dashed vertical line indicates $\lambda = \sqrt{3}$.

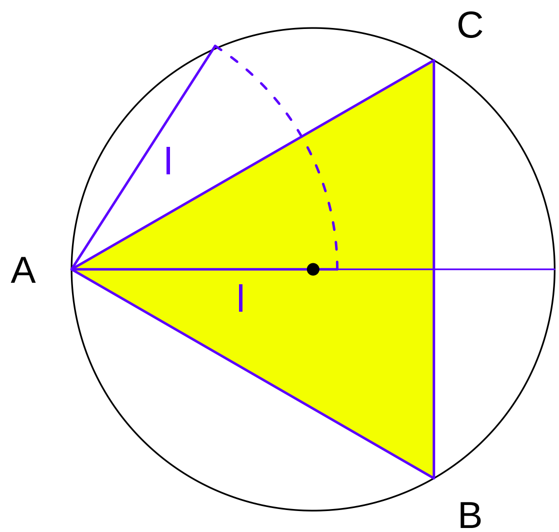

Figure 11: A way to draw chords with lengths uniformly distribute between 0 and $2R$.

2.4 Chords with length uniformly distributed between 0 and $2R$

It is curious that the method of simply taking chords with length uniformly distributed up to the length of the diameter is usually not taken into account, although the idea is not bizarre at all. Indeed this could even be the *natural* procedure to someone used to operate in classical geometry with ruler and compass: place one end of the compass in a point of the circumference (e.g. A in Fig. 11); then place the other end along the diameter impinging the circumference in A; finally rotate the compass in either direction (anticlockwise in the figure) and find the second end of the chord.

The pdf of λ as well as the probabilities of interests are in this case really trivial:

$$f(\lambda\,|\,\mathcal{M}_4) = \frac{1}{2} \qquad (0 \le \lambda \le 2) \tag{35}$$

$$F(\lambda\,|\,\mathcal{M}_4) = \frac{\lambda}{2} \tag{36}$$

$$F(\sqrt{3}\,|\,\mathcal{M}_4) = \frac{\boldsymbol{\sqrt{3}}}{\mathbf{2}} \tag{37}$$

$$F(1\,|\,\mathcal{M}_4) = \frac{1}{2}. \tag{38}$$

Nevertheless, although the results from this extraction model are very easy, if we

	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
$f(\lambda)$	$\frac{1}{\pi\sqrt{1-(\lambda/2)^2}}$	$\frac{\lambda}{4\sqrt{1-(\lambda/2)^2}}$	$\lambda/2$	$1/2$
$F(\lambda)$	$\frac{2}{\pi}\cdot\arcsin\frac{\lambda}{2}$	$1-\sqrt{1-(\lambda/2)^2}$	$\lambda^2/4$	$\lambda/2$
$\mathrm{E}(\lambda)$	$4/\pi \approx 1.27$	$\pi/2 \approx 1.57$	$4/3 \approx 1.33$	1
$\sigma(\lambda)$	$\sqrt{2-16/\pi^2} \approx 0.62$	$\sqrt{8/3-\pi^2/4} \approx 0.45$	$\sqrt{2}/3 \approx 0.47$	$1/\sqrt{3} \approx 0.58$
$P(\lambda \le \sqrt{3})$	$2/3 \approx 0.67$	$1/2 = 0.5$	$3/4 = 0.75$	$\sqrt{3}/2 \approx 0.87$
$P(\lambda \le 1)$	$1/3 \approx 0.33$	$1-\sqrt{3}/2 \approx 0.13$	$1/4 = 0.35$	$1/2 = 0.5$

Table 1: Summary of the four basic chord generators.

ask someone to make a computer program to draw cords 'at random', I would not bet 8.7 to 1.3 (that is $\approx \sqrt{3}/2$ to $1-\sqrt{3}/2$) that a chord will be smaller than the side of the triangle! This will be the subject of section 4.

The results of the four basic methods are summarized in Tab. 1 and in Fig. 12. More details, obtained by simulations, are shown in Figs. 4-7.

3 A chord length generator (and its implementation in R)

In reality, if we are just interested in making a programs with generates the length of the chords 'at random', using one of the four methods we have seen, there is no need to go through all the steps of the operational descriptions of the algorithms. We can just use the probability distributions, summarized in table 1. In order to do that we need a premise, highlighted in the following subsection.

3.1 A curious transformation and its practical importance

Imagine to have a generic continuous variable X whose uncertainty is described by the pdf $f_X(x)$ and the cumulative distribution $F_X(x)$ (in this subsection we shall

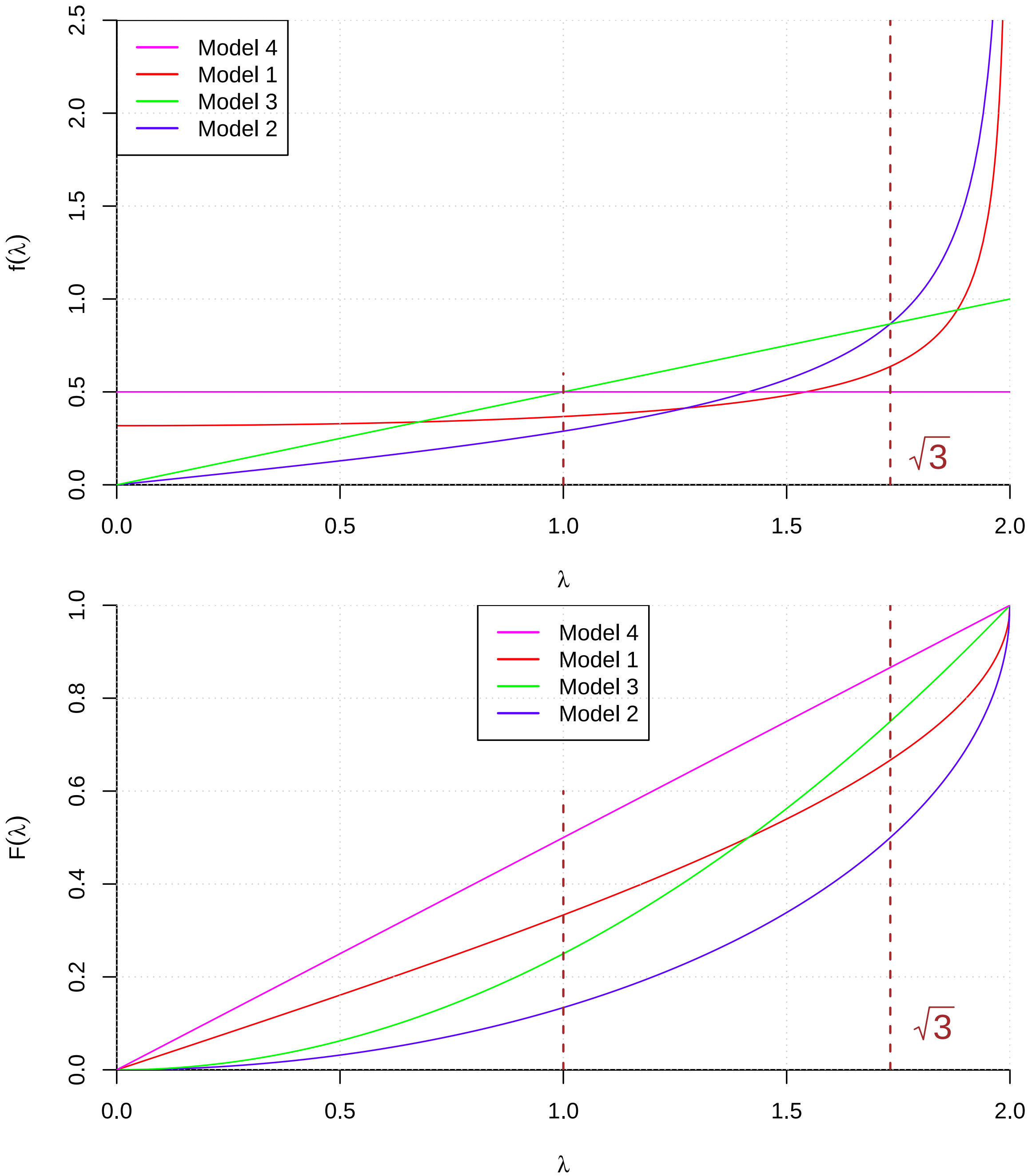

Figure 12: Probability density functions and cumulative function of the lengths of the chords in units of the radius of the circle for the four 'basic' extraction models. (The order of the models in the legend corresponds to the decreasing order of $F(1\,|\,\mathcal{M}_i)$ and $f(0.1\,|\,\mathcal{M}_i)$.)

use the more precise notation introduced in footnote 3). We might think at a new variable Y, related to X by $Y = F_X(X)$, that is

$$y = g(x) = F_X(x)\,, \tag{39}$$

written a a way to stress that in this case $F_X()$ plays now the role of the generic mathematical function $g()$, independently of its probabilistic meaning.

Making use of Eq. 4 we have

$$f_Y(y) = \int_{-\infty}^{+\infty} \delta\left(y - F_X(x)\right) \cdot f_X(x)\, dx\,. \tag{40}$$

Being $F_X()$ monotonic and not-decreasing with derivative equal to $f_x()$, we have then

$$f_Y(y) = \int_{-\infty}^{+\infty} \frac{\delta(x - x^*)}{F'_x(x^*)} \cdot f_X(x)\, dx = \frac{f_X(x^*)}{f_X(x^*)} = 1\,. \tag{41}$$

This is a **great result**, that we double check following the same reasoning used in footnote 3:

$$\begin{aligned} F_Y(y) \equiv P(Y \le y) &= P(Y \le F_X(x)) = P(X \le F_X^{-1}(y)) \equiv F_X\left(F_X^{-1}(y)\right) = y \\ f_Y(y) \equiv \frac{d}{dy} F_Y(y) &= \frac{d}{dy}\, y = 1\,. \end{aligned}$$

Independently of the pdf $f_X()$, the variable Y defined in this way is uniformly distributed between 0 and 1. It follows that the pdf of a variable defined as $X = F_X^{-1}(Y)$ will be $f_x(x)$.

This observation suggests a simple algorithm, very useful in those cases in which the cumulative function is *easy to invert*, to make a (pseudo) random number generator to produce numbers such that our confidence on the occurrence of $X = x$ is proportional to $f(x)$:

$$x = F_X^{-1}(u)\,, \tag{42}$$

where u stands for the occurrence of a uniform (pseudo) random generator that gives numbers (apparently) 'at random' between 0 and 1 (random number generators of this kind are available in all computational environments.)

3.2 Application to the chord problem

Fortunately we can apply this trick to all probability distributions of the chords found in the previous sections. The generic rule (42) gets then implemented as follows.

$\mathcal{M}_1$: $\lambda = 2\sin(\pi u/2)$.

$\mathcal{M}_2$: $\lambda = 2\sqrt{2u - u^2}$.

$\mathcal{M}_3$: $\lambda = 2\sqrt{u}$.

$\mathcal{M}_4$: $\lambda = 2u$.

And this is then, for example, the resulting function written in the R language [6]:

```
rlchords <- function(n, meth) {
  u <- runif(n)
  switch(meth,
         2*sin(pi*u/2),
         2*sqrt(2*u-u^2),
         2*sqrt(u),
         2*u)
}
```

Issuing then the following instruction from an R console ('>' stands for the prompt) you can finally get an histogram similar to the left top one of Fig. 5:

```
> lambda=rlchords(100000,1); hist(lambda, nc=200, col='red')
```

Or you can evaluate mean, standard deviation and fraction of occurrences with λ smaller than $\sqrt{3}$:

```
> mean(lambda)
> sd(lambda)
> length(lambda[lambda<sqrt(3)])/length(lambda)
```

4 What should one expect from a computer drawing program?

"Mater artium necessitas"
("Necessity is the mother of invention")

The original Bertrand problem is about **drawing** chords and not just telling numbers between 0 and 2 (taking a unitary radius). Therefore the question we have to ask is really to draw the chord, by hand of by a computer program.[8] In the introduction I have told what I more or less expect when I ask the practical question, providing a sheet of paper with a pre-designed circle, and I must say that in the last years I had no surprises that induced me to change the model I formed in my mind. You may form yours with practice.

More recently I have also asked PhD students to write "chords generators" with their preferred computer language. As it easy to guess, the choice goes to the algorithm easier to implement, which for physics students is Method 1, since they are familiar with circular motion and with transformations from polar to Cartesian coordinates.

The other methods are somehow tedious because, if taken *literally*, they require several steps with formulae not used everyday, that one needs to derive. For example Method 2, requires *literally*: 1) to choose a radius at random; 2) to chose a point on the radius; 3) find the equation of the line orthogonal to the radius in that point; 4) find the interceptions of the line with the line with the circle. And – I must confess, with some shame – that the first time I was playing with the problem, I was implementing in R these detailed procedures. When during this year course I tried to make an Android app to draw 'random' chords on a circle, using App Inventor [7], I was horrified by the formulae I had to 'write' with that tool. So I initially implemented only Method 1. After a while I also implemented Method 2, but not following the procedures described above. The trick was to extract a point along the horizontal diameter, thus coinciding with the abscissa and operate then a random rotation. I that way I was able reuse somehow the 'blocks' (the graphical programming elements shown e.g. in Fig. 13) developed for Method 1.

The rest of this section is devoted to simulation issues, showing how to avoid pedantic procedures and without pretending that the suggested algorithms are the 'best' in some sense that should be better defined.

[8]Nevertheless, the lengths provided by the 'chord generator' presented in the previous section do provide valid answers, since the pdf's have been derived using some geometric rules to produce chords and not with an abstract algorithm to produce numbers between 0 and 2, like those you would get e.g. with the following R command

```
> n=10; lambda = 2*sin(runif(n, 0, pi))^2
```

4.1 Model 1 ($\mathcal{M}_1$)

As stated above, this is the one that appears the simplest (to implement in a program) to physics students, to most colleagues and to myself. Here is how it appears in R (`n` is the number of chords).

```
> ph1 <- runif(n, 0, 2*pi)
> ph2 <- runif(n, 0, 2*pi)
> p1 <- cbind(cos(ph1), sin(ph1))
> p2 <- cbind(cos(ph2), sin(ph2)
> l <- sqrt( (p2[,1]-p1[,1])^2 + (p2[,2]-p1[,2])^2)
```

The result is a 'vector' of `n` lengths `l` (in units of the radius). Plus we have the matrices of interception points (each raw is a point).

4.2 Model 2 ($\mathcal{M}_2$)

In this case we start extracting x between -1 and 1, evaluating the corresponding ordinates.

```
> p1 <- p2 <- runif(n, -1, 1)
> p1 <- cbind(p1, sqrt(1-p1^2))
> p2 <- cbind(p2, -sqrt(1-p2^2))
```

Then we define a random rotation angle and add it to the polar angles calculated from the points:

```
> phr <- runif(n, 0, pi)
> ph1 <- phr + atan2(p1[,2],p1[,1])
> ph2 <- phr + atan2(p2[,2],p2[,1])
```

At this point we have reuse exactly the last three lines of code of the previous method. The resulting App Inventor blocks are shown in Fig. 13.

4.3 Model 3 ($\mathcal{M}_3$)

To chose a point uniformly inside the circle we could extract uniformly x and y between -1 and 1 and discard the points which are outside the circle of radius 1. But we can use of the previous code (or App Inventor blocks) if we extract a point along the radius with pdf $f(\rho) = 2\rho$, as we have learned in subsection 2.3, making use of the technique learned in subsection 3.1.

```
> rho <- sqrt(runif(n))
```

But, to reuse the previous code, we have to invert at random (with probability 1/2) the sign of this numbers. Technically this cab be done in R creating a vector of random -1 and 1 obtained by a binomial generator and multiplying element by element. Thus our starting abscissas will be

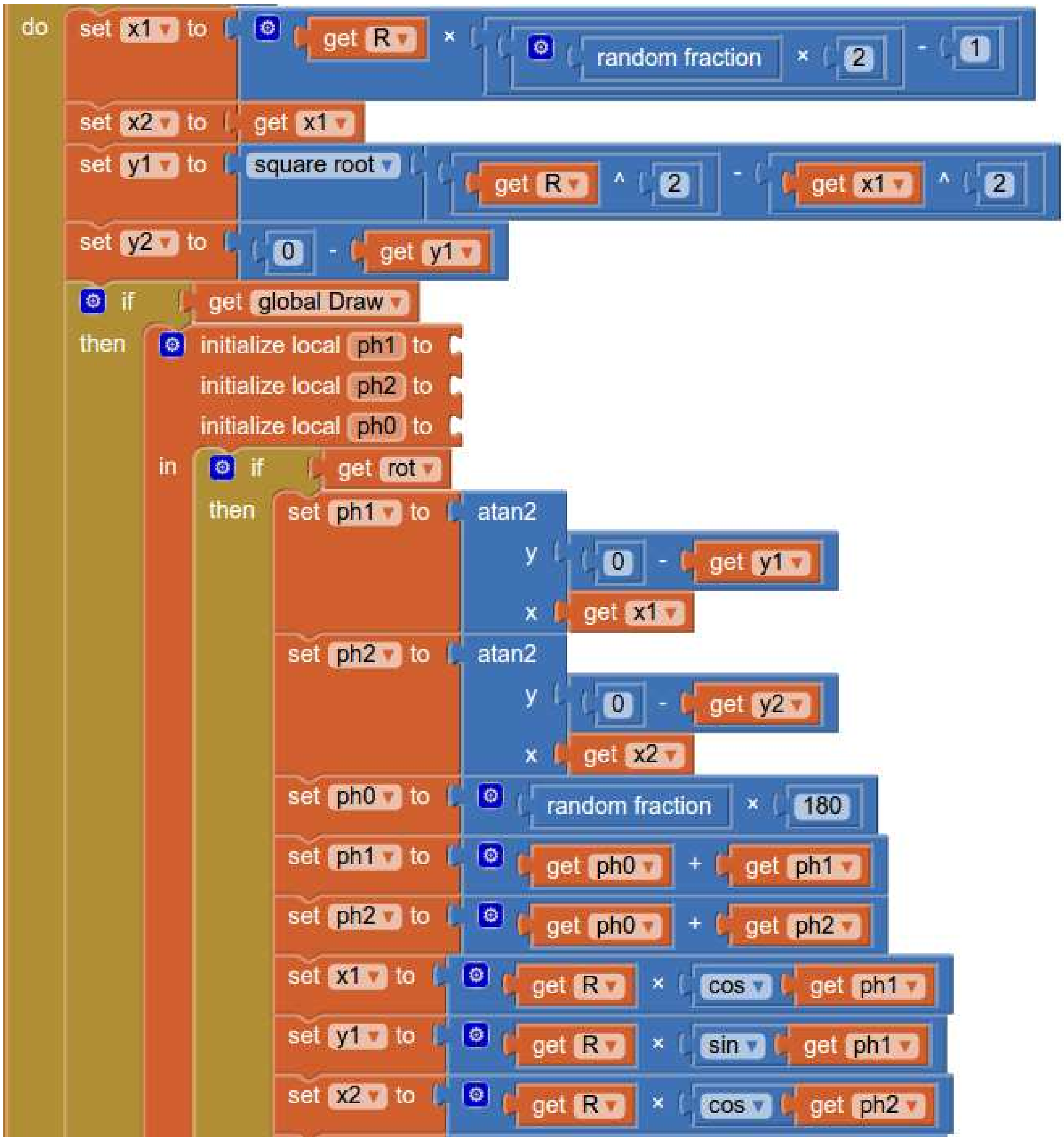

Figure 13: App Inventor blocks for the core of Method 2 (see text).

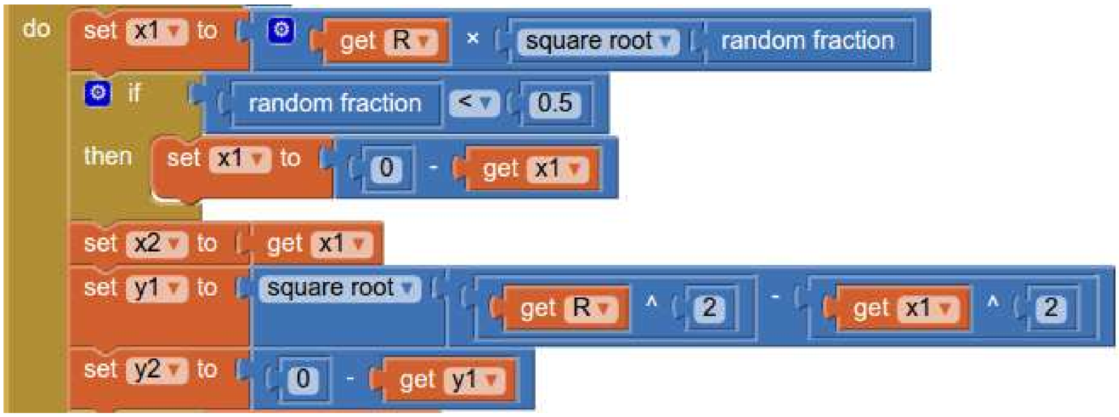

Figure 14: App Inventor blocks to turn Method 2 into Method 3 (see text).

```
> p1 <- p2 <- rho * ( rbinom(n, 1, 0.5)*2 - 1 )
```

After this, we continue exactly as in the second line of R code of the previous method. The implementation of this variation in App Inventor is shown in Fig. 14.

4.4 Model 4 ($\mathcal{M}_4$)

Also in the case of the fourth method, we can reuse the code written for Method 2, without having to calculate the intersections of two circles. With the help of Fig. 15 we can recognize a useful rectangular triangle and then make use of a famous theorem of elementary geometry. The projection p is then $l^2/2R$ and then the abscissa of the intersection is equal to $x = p - R$. Here is how to recover, once more, the code of Method 1 (remember that we use unitary radius):

```
> p1 <- p2 <- runif(n, 0, 1)^2 - 1
```

and all the rest we be the same.

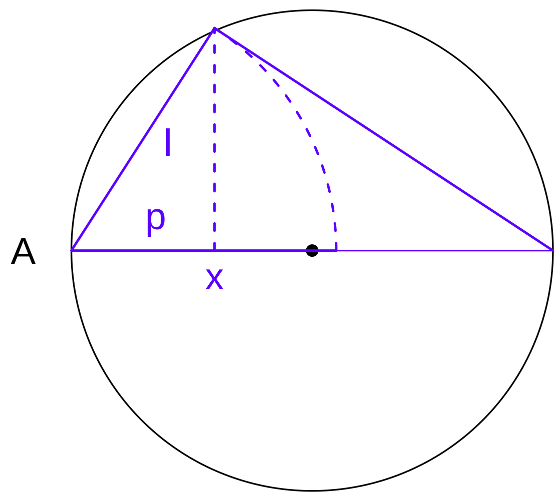

Figure 15: A sketch which show how to calculate the abscissa of the the interception (x) by simple geometry (see text).

5 A physical model: throwing mikado sticks on a pattern of circles

The authors of Ref. [5] claim to have found a 'conclusive physical solution' to the Bertrand problem, but I have strong doubts that they have ever tried to implement their model in a real experiment. Something which seems to me more realistic is the kind of game sketched in Fig. 16: a pattern of circles[9] on a table, or on the floor, on which we throw mikado sticks, or toothpicks, needles or something similar. The only, obvious, conditions is on the minimum length of the sticks, that has to be at least the double of the maximum diameter of the circles.[10]

If we throw 'ad random' the sticks, somehow towards the center of the pattern in order to avoid complications with boundary conditions, we expect their centers and their orientations 'uniformly' distributed (the former in the plane, the latter in angle w.r.t. a given direction). We consider only sticks whose reference point, marked somehow, is inside one circle and consider the resulting chords defined by the

[9]In Fig. 16 all circles have the same size, but this is not a necessary requirement, as it will be clear in a while.

[10]Also this condition is not necessary, if we think to prolong the stick in either end by a ruler to draw the chords. And, finally, it is not even required that the reference point, needed to decide inside which circle the chord has to be drawn, has to coincide with the center of the stick. The formulation in the text is, or at least so seems to me, the easiest to be implemented in a real 'game', similar to the famous "Buffon's needle".

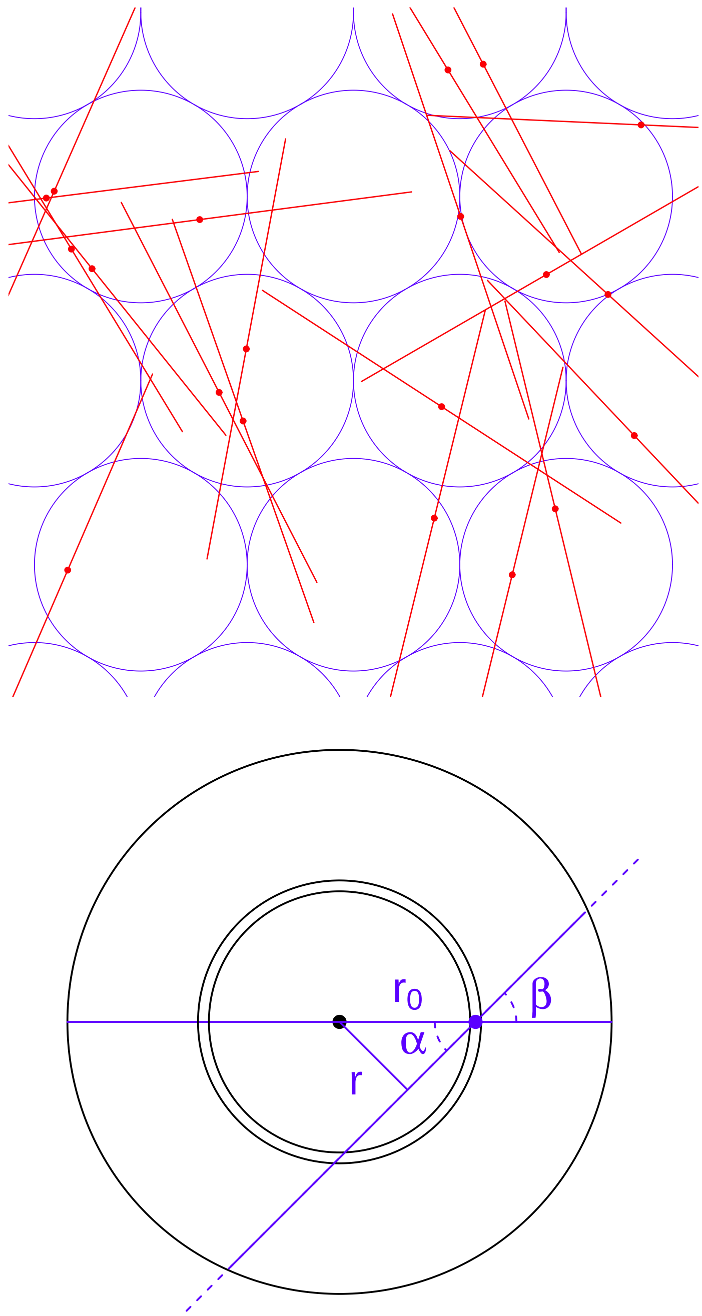

Figure 16: Above: Mikado sticks thrown at random on a pattern of circles. Below: Sketch of the mikado experiment to show how to evaluate the distance of the chord from the center from the position of the center of the stick and its orientation (see text).

intersection of the stick with the circumference of that circle. As we can see from the figure, the efficiency is rather high (and, as a byproduct, we can try to estimate empirically the area of the regions between circles, but this is another story…).

Evaluating the pdf of the expected chords length might be complicated, but fortunately we can make use of some of our previous results. Let us indicate now with r_0 the distance from the center of the stick to the center of the circle and with r the distance of the chord from the center of the circle (see Fig. 16). The respective quantities in units of R are then $\rho_0 = r_0/R$ and $\rho = r/R$.

By the hypothesis inherent to the throwing mechanism the center of the stick is uniformly distributed in the circle. This implies, as we already know, that high values of ρ_0 are more probable than small ones, namely $f(\rho_0\,|\,\mathcal{M}_5) = 2\rho_0$. The fractional distance r is related to r_0 by $r = r_0 \sin\alpha$, where the angle α, defined in the construction in Fig. 16, is uniformly distributed between 0 and $\pi/2$. These are then our starting conditions [11]

$$\rho = \rho_0 \sin\alpha \tag{43}$$

$$f(\rho_0\,|\,\mathcal{M}_5) = 2\rho_0 \qquad (0 \le \rho_0 \le 1) \tag{44}$$

$$f(\alpha\,|\,\mathcal{M}_5) = \frac{1}{\pi/2} = \frac{2}{\pi} \qquad (0 \le \alpha \le \frac{\pi}{2})\,. \tag{45}$$

We can then calculate the probability distribution of ρ and make then use of the reasoning applied in Method 3 to evaluate the probability distribution of λ. Indeed, for a given ρ_0, the pdf of ρ, conditioned by the value of ρ_0, will be given by

$$f(\rho\,|\,\mathcal{M}_5, \rho_0) = \int_0^{\pi/2} \delta(\rho - \rho_0\,\sin\alpha)\cdot f(\alpha)\,d\alpha \tag{46}$$

$$= \int_0^{\pi/2} \frac{\delta(\alpha-\alpha^*)}{\rho_0\,\cos\alpha^*}\cdot\frac{2}{\pi}\,d\alpha\,, \tag{47}$$

with $\alpha^* = \arcsin(\rho/\rho_0)$, from which it follows

$$f(\rho\,|\,\mathcal{M}_5, \rho_0) = \frac{2}{\pi\,\rho_0\sqrt{1-(\rho/\rho_0)^2}} \tag{48}$$

[11] An alternative way would be to use the angle β defined in Fig. 16, ranging between 0 to π, with $f(\beta\,|\,\mathcal{M}_5) = 1/\pi$. The angle α inside the rectangular triangle will be equal to β if β is smaller than $\pi/2$, and $\pi-\beta$ elsewhere. The relation (43) would then be replaced by $\rho = \rho_0\,\sin\beta$ and the integral (46) replaced on the equivalent one in $d\beta$ between 0 and π, with the factor $2/\pi$ in the integrand replaced by $1/\pi$, *apparently* leading to results differing by a factor of 2. This apparent contradiction is resolved noted that transformation rule of the Dirac delta has now two roots, $\beta_1^* = \arcsin(\rho/\rho_0)$ and $\beta_2^* = \pi - \arcsin(\rho/\rho_0)$. But, since $|\cos\beta_1^*| = |\cos\beta_2^*|$, we have two identical contributions, thus exactly compensating the missing factor 2.

$$= \frac{2}{\pi\sqrt{\rho_0^2-\rho^2}}\,. \tag{49}$$

Having $f(\rho\,|\,\mathcal{M}_5,\rho_0)$ and $f(\rho_0\,|\,\mathcal{M}_5)$ we can then evaluate $f(\rho\,|\,\mathcal{M}_5)$ as

$$f(\rho\,|\,\mathcal{M}_5) = \int_\rho^1 f(\rho\,|\,\mathcal{M}_5,\rho_0)\cdot f(\rho_0\,|\,\mathcal{M}_5)\,d\rho_0\,, \tag{50}$$

in which we have to pay attention to the condition $\rho \le \rho_0$. The result is

$$f(\rho\,|\,\mathcal{M}_5) = \int_\rho^1 \frac{2}{\pi\sqrt{\rho_0-\rho^2}}\cdot 2\rho_0\,d\rho_0 \tag{51}$$

$$= \int_\rho^1 \frac{4\,\rho_0}{\pi\sqrt{\rho_0^2-\rho^2}}\,d\rho_0 \tag{52}$$

$$= \frac{4}{\pi}\sqrt{1-\rho^2}\,. \tag{53}$$

Having calculated the pdf of the distances of the chords from the center of the circle, we continue as in Eq. (27), thus obtaining

$$f(\lambda\,|\,\mathcal{M}_5) = \int_0^1 \delta\left(\lambda - 2\sqrt{1-\rho^2}\right)\cdot\frac{4}{\pi}\sqrt{1-\rho^2}\,d\rho \tag{54}$$

$$= \int_0^1 \frac{\delta\,(\rho-\rho^*)}{2\rho^2/\sqrt{1-\rho^2}\Big|_{\rho=\rho^*}}\cdot\frac{4}{\pi}\sqrt{1-\rho^2}\,d\rho \tag{55}$$

$$= \frac{4/\pi\,\sqrt{1-\rho^{*2}}}{2\rho^*/\sqrt{1-\rho^{*2}}} = \frac{2}{\pi}\,\frac{1-\rho^{*2}}{\rho^*}\,, \tag{56}$$

with the usual $\rho^* = \sqrt{1-(\lambda/2)^2}$. We get finally

$$f(\lambda\,|\,\mathcal{M}_5) = \frac{2}{\pi}\,\frac{(\lambda/2)^2}{\sqrt{1-(\lambda/2)^2}} \tag{57}$$

$$F(\lambda\,|\,\mathcal{M}_5) = \frac{2}{\pi}\arcsin(\lambda/2) - \frac{\lambda}{\pi}\sqrt{1-(\lambda/2)^2}\,, \tag{58}$$

shown in Fig. 17 and from which we can calculate the usual indicators

$$\mathrm{E}(\lambda\,|\,\mathcal{M}_5) = \frac{16}{3\,\pi} \approx 1.70 \tag{59}$$

$$\sigma(\lambda\,|\,\mathcal{M}_5) = \sqrt{3-\frac{256}{9\,\pi^2}} \approx 0.34 \tag{60}$$

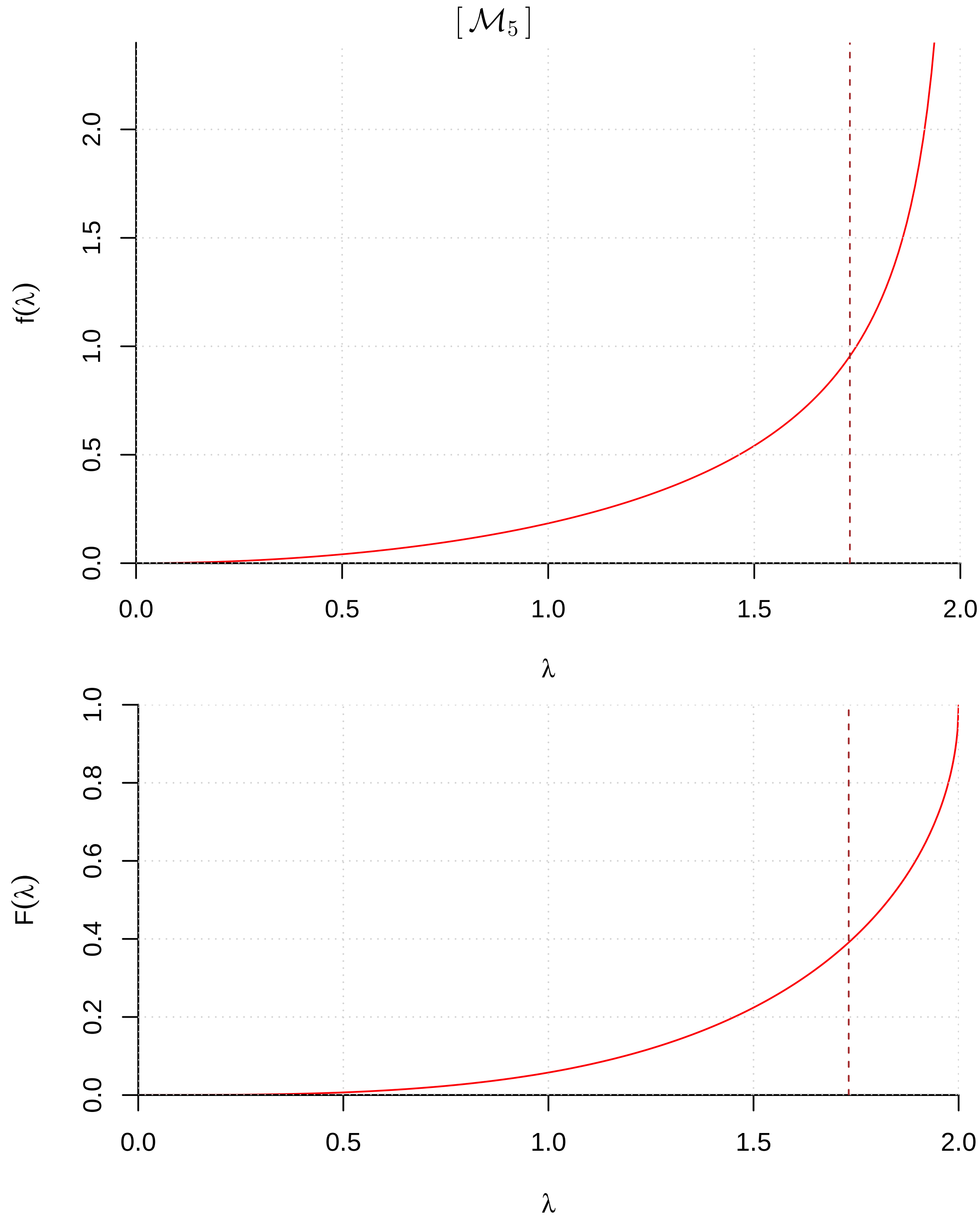

Figure 17: Probability distribution of $\lambda = l/R$ of the chords generated with Method 5 ('mikado'). The dashed vertical line indicates $\lambda = \sqrt{3}$.

$$P(\lambda \leq \sqrt{3}\,|\,\mathcal{M}_5) \;=\; \frac{2}{3} - \frac{\sqrt{3}}{2\,\pi} \approx 0.39 \tag{61}$$

$$P(\lambda \leq 1\,|\,\mathcal{M}_5) \;=\; \frac{1}{3} - \frac{\sqrt{3}}{2\,\pi} \approx 0.058\,. \tag{62}$$

The other details obtained by simulation, reported in Figs. 4-7, for the four 'basic' methods are shown in Fig. 18.

6 Conclusions

The only sound experiment I could think about (but perhaps I miss of fantasy) leads to a solution different than the claimed 'final' solutions of the problem.[4, 5]

References

[1] J. Bertrand, *Calcul Des Probabilites*, 1907, `http://www.archive.org/details/calculdesprobabi028601mbp`, pp. 4-5.

[2] `http://en.wikipedia.org/wiki/Bertrand_paradox_%28probability%29`

[3] E. Schrödinger, *The foundation of the theory of probability - I*, Proc. R. Irish Acad. 51A (1947) 51.

[4] E.T. Jaynes, *The Well-Posed Problem*, Foundations of Physics 3: 477–493 (1973).

[5] P. Di Porto et al., *Bertrand's paradox: a physical solution*, arXiv:1008.1878.

[6] R

[7] AI2

$[\mathcal{M}_5]$

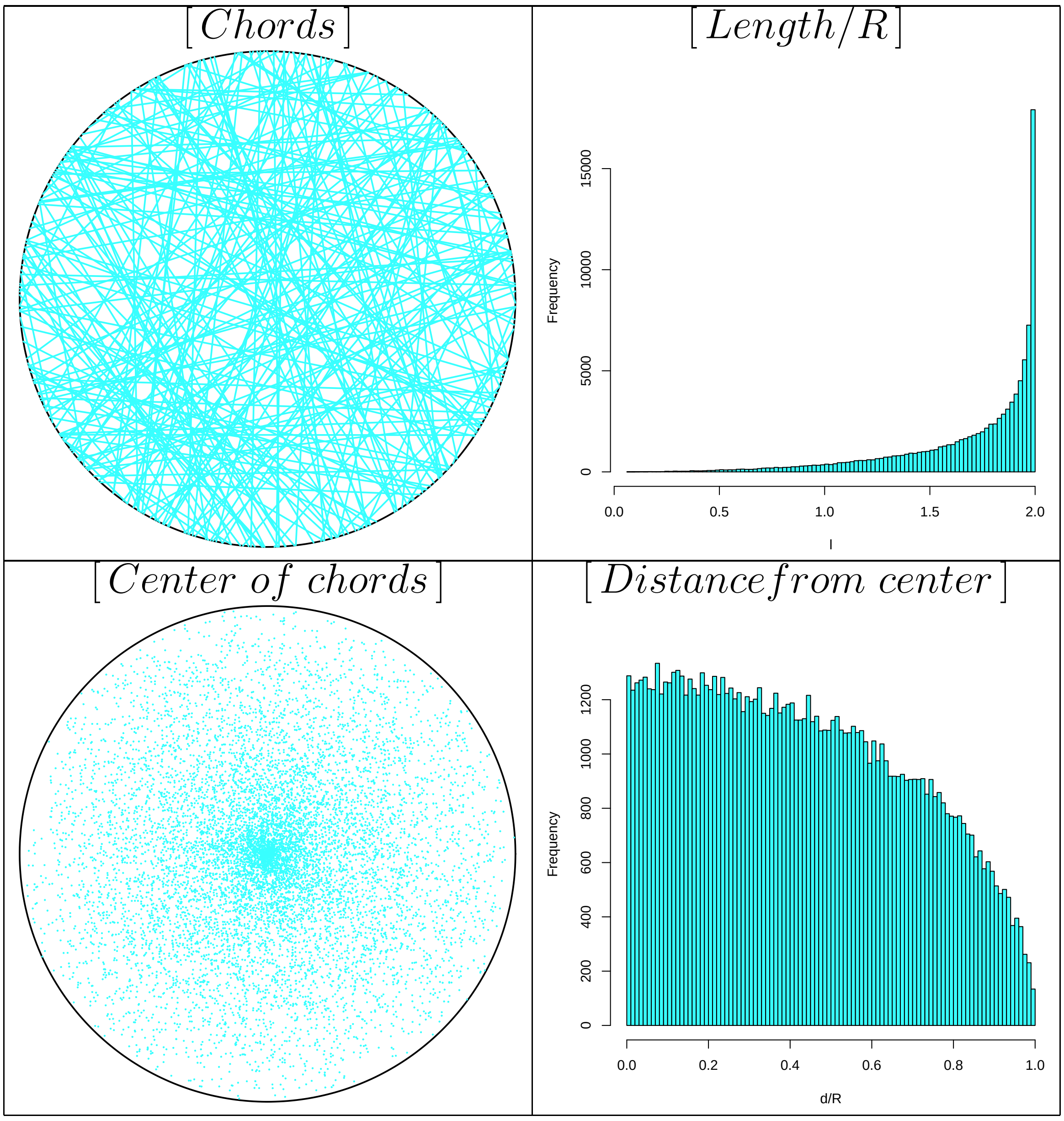

Figure 18: Results from simulations based on the ‘mikado model’, or Model 5.