Bertrand 'paradox' reloaded
(with details on transformations of variables, an introduction to Monte Carlo simulation and an inferential variation of the problem)[*]

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 mainly 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 and an appendix on the inferential aspects of the problem.)

> *"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. The Android app mentioned in the text is available at `http://www.roma1.infn.it/~dagos/prob+stat.html#bpr`.

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 see.

The question can be restated in more general terms, i.e. that of finding the probability distribution of the 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', $f(l)$. This should be written, more precisely, as $f(l \,|\, I_s(t))$, where $I_s(t)$ 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 the use of 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 experimental setup!).

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 (the bet could be even more detailed, concerning for example the orientation of the chord – for example 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, as we shall see in Appendix C).

Indeed I have done experiments of this kind since several 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 possible <u>real</u> '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 the search engine?"] When instead I propose 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 segment 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. The paper is supplemented by three appendices. Appendix A is a kind of extra exercise on transformation of variables. Appendix B show a simple way to write an 'infinite number' of chord generators (indeed a generator characterized by two *hyperparameters*). Appendix C is finally an inferential variation of Bertrand problem, in which we use 'experimental data' to guess the generator used and to predict the length of other chords produced with the same generator.

2 Basic ('classic') geometric 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. The adjective *geometric* is to distinguish then from a more 'physical' one, based on a kind of realistic game that, as we shall see in section 6.

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 either in the arc between A and B or in that between A and C. The resulting probability is thus simply **2/3**.

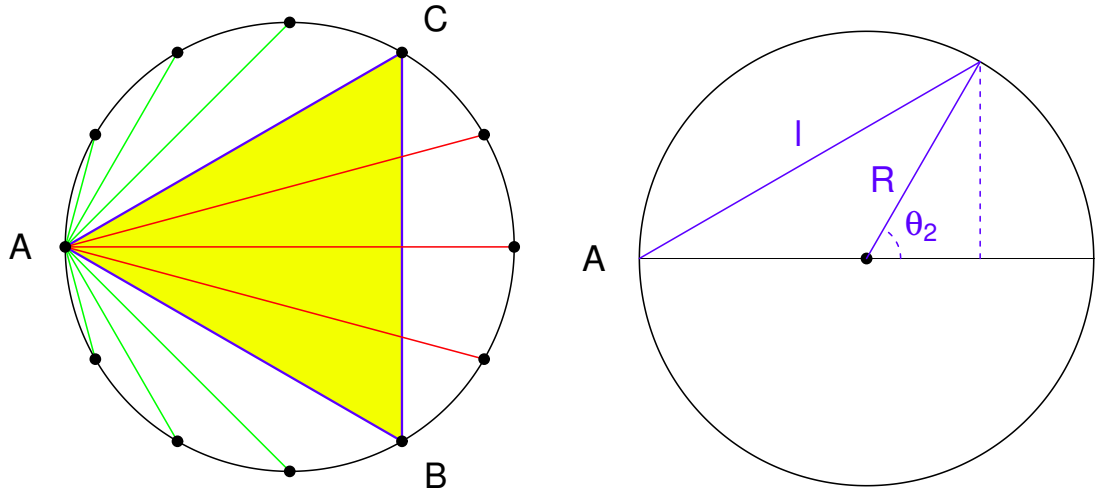[1] `https://www.google.it/search?q=chord+geometry&source=lnms&tbm=isch`

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).

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 one to one 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 extraction method 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)

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

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

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

---

[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.

We shall use the general rule [3]

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

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 infinitesimal probabilities '$f(\theta_2 \,|\, \mathcal{M}_1)\, d\theta_2$' 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 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 $\lambda(\theta_2)$

---

[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:

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

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 $f_Y()$]:

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

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

Making use of the rules of calculus we have then

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

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

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

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.

decreases monotonically we get [4]

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

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

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)

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

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

---

[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

$$f_Y(y) = \int_{-\infty}^\infty \delta(y - g(x)) \cdot f_X(x) \, dx$$

$$= \int_{-\infty}^\infty \frac{\delta(x - x^*)}{|g'(x^*)|} \cdot f_X(x) \, dx$$

$$= \frac{1}{|g'(x^*)|} \cdot f_X(x^*) \,,$$

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.

6

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.

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:

$$P(\lambda \le \sqrt{3} \,|\, \mathcal{M}_1) = F(\sqrt{3} \,|\, \mathcal{M}_1) \;=\; \frac{2}{\pi} \cdot \arcsin\left(\frac{\sqrt{3}}{2}\right) = \frac{2}{\pi} \cdot \frac{\pi}{3} = \mathbf{\frac{2}{3}}$$

$$P(\lambda \le 1 \,|\, \mathcal{M}_1) = F(1 \,|\, \mathcal{M}_1) \;=\; \frac{2}{\pi} \cdot \arcsin\left(\frac{1}{2}\right) = \frac{2}{\pi} \cdot \frac{\pi}{6} = \frac{1}{3}$$

Simulations of chords with this methods are reported in the top left plot of Figures 19-23 at the end of the paper: Fig. 19 shows a sample of random chords; Fig. 21 the position of their center. The distribution of the lengths in units of R are shown in the histograms of Fig. 20, while those of Fig. 22 show the distribution of the distance of the chords from the center of the circle, about which we shall say more in Appendix A. Finally, Fig. 23 is like Fig. 19, but 'without rotation', the meaning of this espression becoming clearer as we go through the various methods. In the case of the first method this corresponds to the way we have followed to derive $f(\lambda \,|\, \mathcal{M}_1)$ in this subsection, that is fixing θ_1 at π and extracting θ_2 between 0 and π.

---

[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

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

also yielding Eq. (12). (The minus sign resulting from the derivation is because λ decreases as θ_2 increases, as clear from Fig 1.)
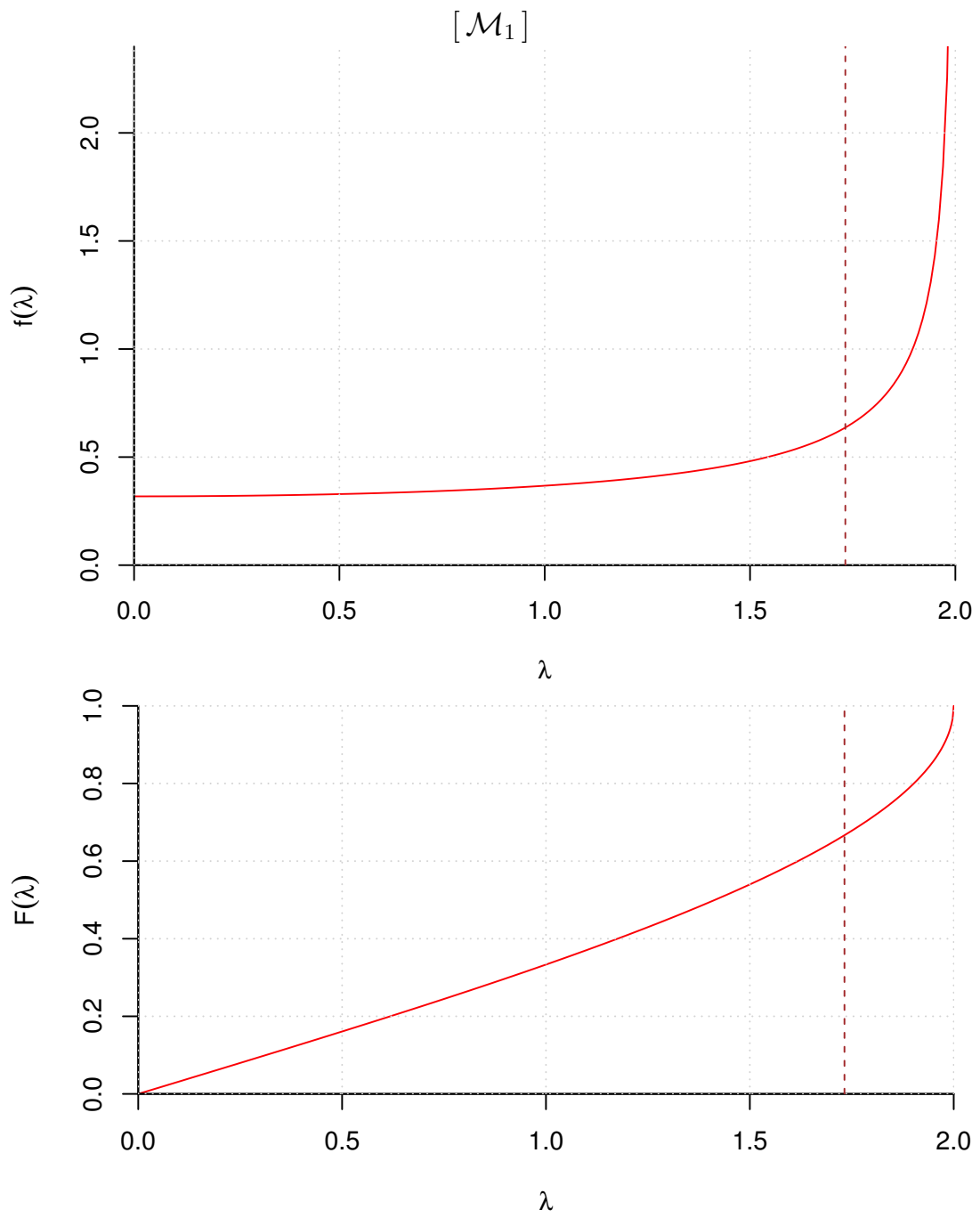
Figure 2: Probability distribution of $\lambda = l/R$ of the chords generated with Method 1. The dashed vertical line indicates $\lambda = \sqrt{3}$.
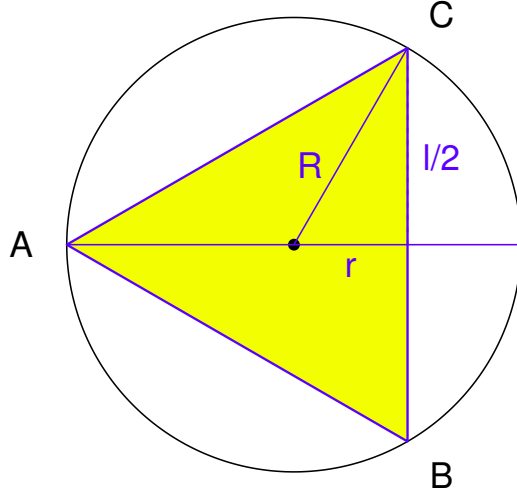
Figure 3: Construction of a chord orthogonal to a radius and distant r from the center of the circle.

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

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

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

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

Then, the pdf of interest will be given by

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

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

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

9

with

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

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

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

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

plotted in Fig. 4 and from which we can calculate the probabilities of interest:

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

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

Simulations of chords with this methods are reported in the plots of Figs. 19-23.

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 centers 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)

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

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. 5). The pdf of ρ is then, after 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

$$f(\lambda \,|\, \mathcal{M}_3) = \int_0^1 \delta\left(\lambda - 2\sqrt{1 - \rho^2}\right) \cdot 2\rho \, d\rho \tag{27}$$

---

[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 $f_P(\rho) = 1$:

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

$$= 1 \cdot \frac{\lambda}{4\sqrt{1 - (\lambda/2)^2}},$$
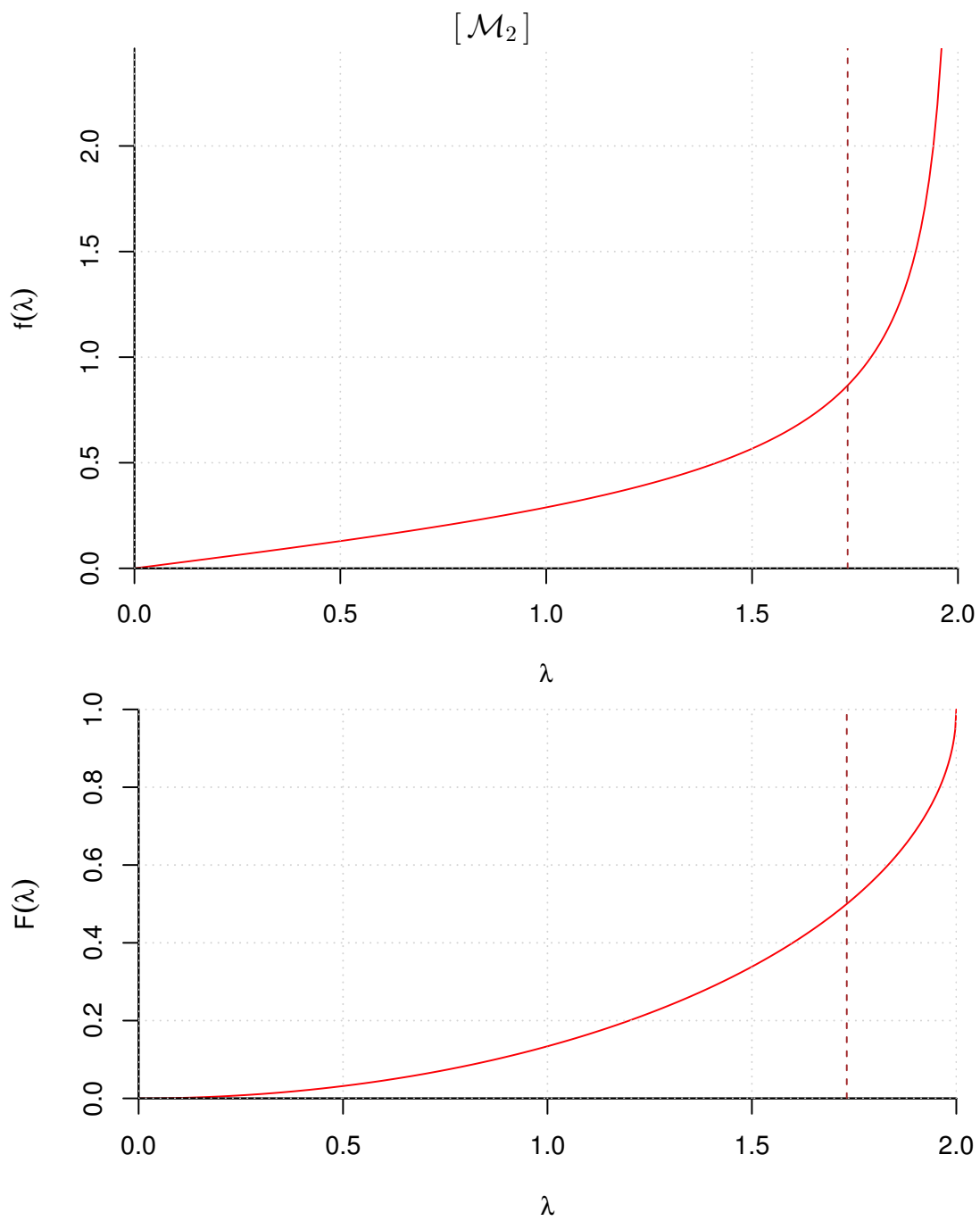
that is precisely Eq. (21).

Figure 4: Probability distribution of $\lambda = l/R$ of the chords generated with Method 2. The dashed vertical line indicates $\lambda = \sqrt{3}$.
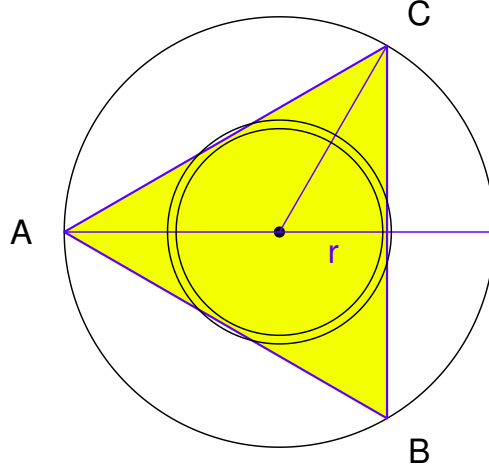
Figure 5: As Fig. 3, but with the annulus of infinitesimal width dr drawn at r to show that in Model 3 the infinitesimal probability dP that the distance of a chord from the center is between r and $r+dr$ is given by $dP \equiv dF(r) = 2\pi r\,dr/(\pi R^2) = (2r/R^2)\,dr$, and hence $dF(\rho) = 2\rho\,d\rho$, or, in our notation, $f(\rho\,|\,\mathcal{M}_3) = 2\rho$, with $\rho = r/R$.

$$= \int_0^1 \frac{\delta(\rho - \rho^*) \cdot 2\rho}{\left| \left(\frac{d}{d\rho}(\lambda - 2\sqrt{1-\rho^2}) \right) \big|_{\rho=\rho^*} \right|}\, d\rho\,, \tag{28}$$

$$= \frac{2\,\rho^*}{2\rho^*/\sqrt{1-\rho^{*2}}} = \sqrt{1-\rho^{*2}} \tag{29}$$

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

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

The result can be then summarized as

$$f(\lambda\,|\,\mathcal{M}_3) = \frac{\lambda}{2} \qquad (0 \le \lambda \le 2) \tag{31}$$

$$F(\lambda\,|\,\mathcal{M}_3) = \frac{\lambda^2}{4}\,, \tag{32}$$

from which we obtain

$$F(\sqrt{3}\,|\,\mathcal{M}_3) = \frac{\mathbf{3}}{\mathbf{4}} \tag{33}$$

$$F(1\,|\,\mathcal{M}_3) = \frac{1}{4}. \tag{34}$$

---

[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:

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

$$= 2\sqrt{1-(\lambda/2)^2} \cdot \frac{\lambda}{4\sqrt{1-(\lambda/2)^2}} = \frac{\lambda}{2}\,.$$
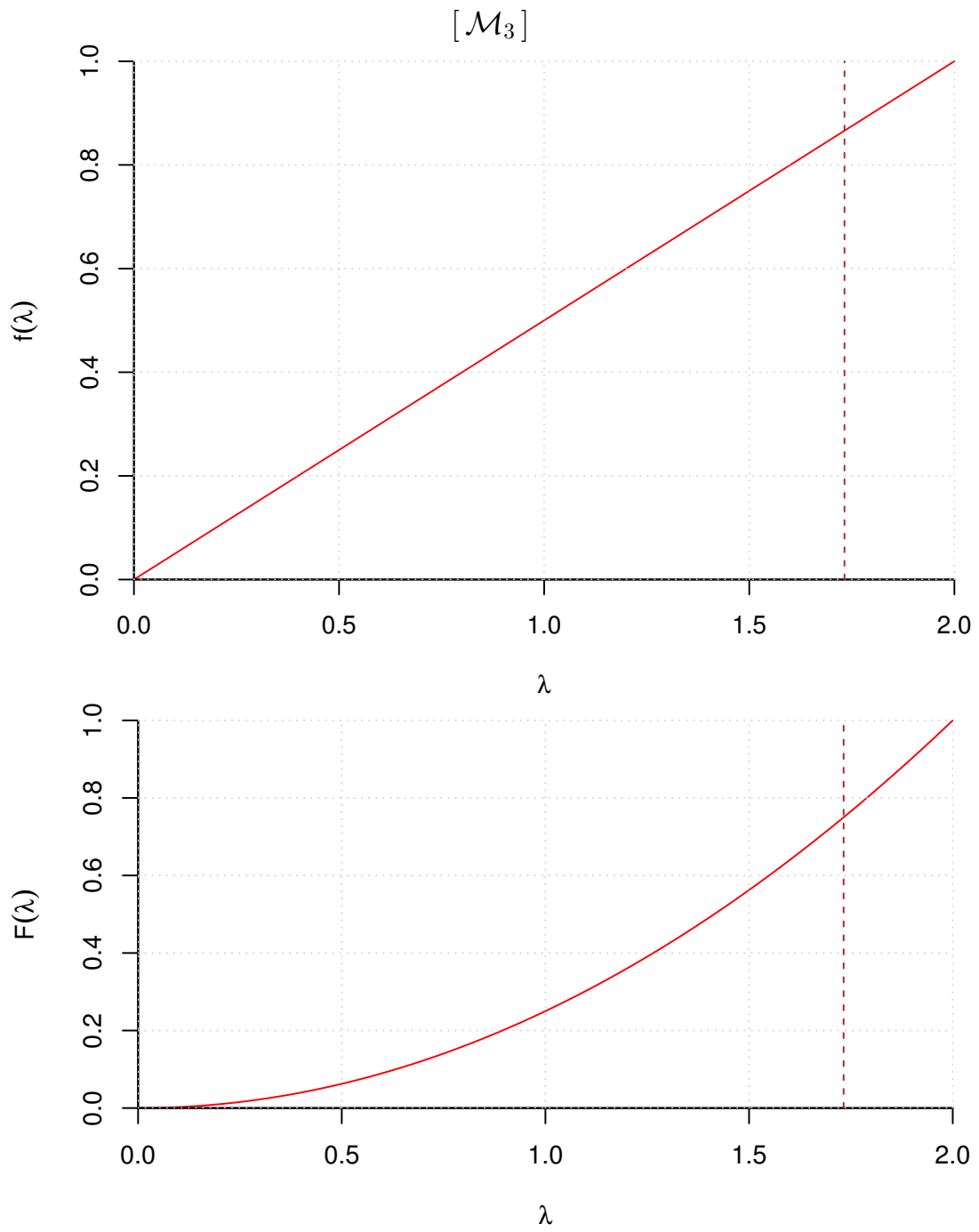
$$[\mathcal{M}_3]$$

Figure 6: Probability distribution of $\lambda = l/R$ of the chords generated with Method 3. The dashed vertical line indicates $\lambda = \sqrt{3}$.
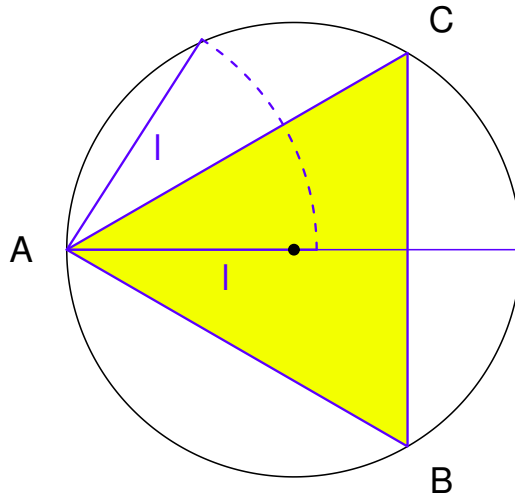
Figure 7: A way to draw chords with lengths uniformly distribute between 0 and $2R$.

Simulations of chords with this methods are reported in the plots of Figs. 19-23.

3   Some "ruler and compass" methods

Let us now see another two 'geometric' methods in which chords a drawn by opening at random an ideal compass ('ideal' because the minimum distance between the end points is taken to be zero) from a reference point along the circumference. (By 'opening at random' we mean that the endpoints of the compass will define, uniformly, segments up to the diameter. Changing the opening method we can then define two classes of chord generators.)

3.1   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 the needle of the compass in a point of the circumference (e.g. A in Fig. 7) to define one endpoint of the chord; then place the pencil lead along the diameter impinging the circumference in A; finally rotate the compass in either direction (anticlockwise in the figure) and find the second endpoint of the chord.

The probability distribution of λ (see Fig. 8) as well as the probabilities of interests are in this case really trivial:

$$f(\lambda \,|\, \mathcal{M}_4) \;\; = \;\; \frac{1}{2} \quad (0 \leq \lambda \leq 2) \tag{35}$$
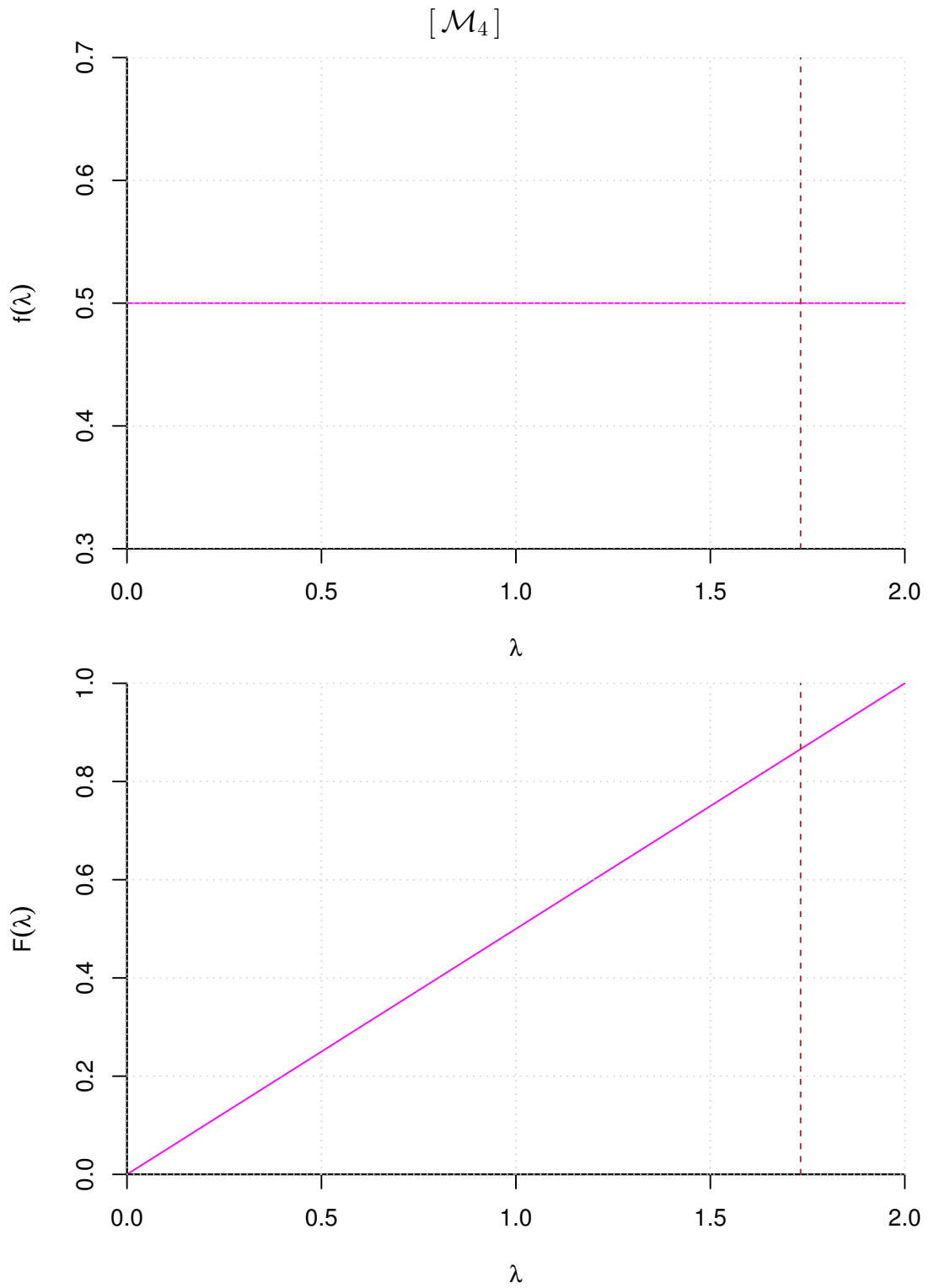
14

Figure 8: Probability distribution of $\lambda = l/R$ of the chords generated with Method 4. The dashed vertical line indicates $\lambda = \sqrt{3}$.
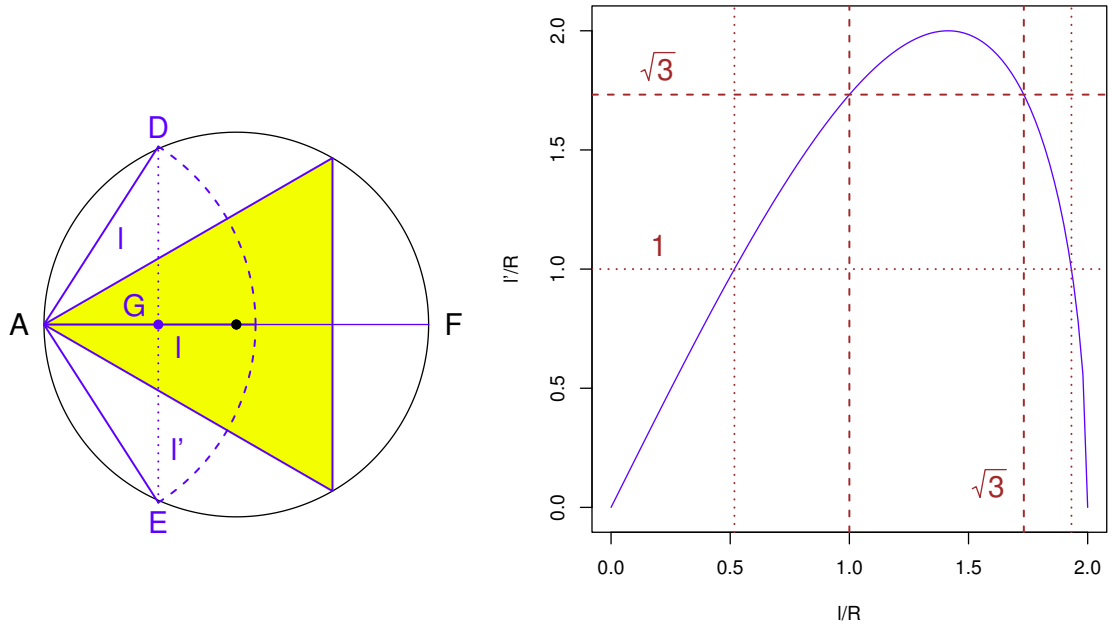
Figure 9: An alternative way to drawn chords with ruler and compass. Left: construction of the chord. Right: relation between the l and l' (both in units of the radius).

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

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

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

Simulations of chords with this methods are reported in the plots of Figs. 19-23.

Nevertheless, although the results from this extraction model are very easy, if we ask someone to make a computer program to really <u>draw</u> 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 point will become clear in section 5.

3.2   A variant of Method 4

When using ruler and compass it is almost *automatic* that, after having drawn an arc in one direction to intercept the circumference, one rotates the tool the other direction, thus identifying the two points E and D of Fig. 9, which then become the *natural* endpoints of a chord (the reader understands that at this point the use of the adjective 'natural' is at limit of being sarcastic). The length l' of this chord is related to the segment l defined[8] by the opening of the compass by $l' = 2\,l\,\sqrt{1 - (l/2R)^2}$, or,

---

[8]The points ADF define a rectangular triangle. The length of the segment $\overline{AG}$ is then equal to $l^2/2R$, while the square of the length of $\overline{DG}$ is equal to the product of the lengths of $\overline{AG}$ and of

in units of the radius,

$$\lambda' = 2\lambda\sqrt{1 - (\lambda/2)^2}. \tag{39}$$

This relation is shown in the right plot of Fig. 9, from which we can calculate the usual probability that λ' is smaller that $\sqrt{3}$. This is equal to the probability that l is smaller than 1, that is $1/2$, plus the probability that it is larger that $\sqrt{3}$, that is $(2 - \sqrt{3})/2$. The result is then $(\mathbf{3} - \sqrt{\mathbf{3}})/\mathbf{2} \approx 63.4\%$, that do not correspond to none of the previous methods. Similarly, the probability that the chord is smaller than the radius, or that $\lambda' < 1$, is equal to the probability that λ is smaller than $\sqrt{2 - \sqrt{3}}$, plus the probability that it is larger than $\sqrt{2 + \sqrt{3}}$, that is [9]

$$\frac{\sqrt{2 - \sqrt{3}}}{2} + \frac{2 - \sqrt{2 + \sqrt{3}}}{2} = 1 - \frac{1}{\sqrt{2}} \approx 29.3\,\%. \tag{40}$$

(This second calculation is important in order to double check the result that we shall get later, being the resulting formulae not very 'pretty'.)

Also in this case it is possible to arrive to a closed expression of the pdf. We do here the exercise mainly to show a case in which the property of the Dirac delta needed to make the transformation involves more than one root, as evident from the non-monotonic relation shown in Fig. 9. Here are the details:

$$\begin{aligned}
f(\lambda'\,|\,\mathcal{M}_5) &= \int_0^2 \delta\left(\lambda' - 2\lambda\sqrt{1 - (\lambda/2)^2}\right) \cdot \frac{1}{2}\, d\lambda \\
&= \int_0^2 \left(\frac{\delta(\lambda - \lambda_1^*)}{|\frac{d}{d\lambda}(-2\lambda\sqrt{1 - (\lambda/2)^2})|_{\lambda=\lambda_1^*}} + \frac{\delta(\lambda - \lambda_2^*)}{|\frac{d}{d\lambda}(-2\lambda\sqrt{1 - (\lambda/2)^2})|_{\lambda=\lambda_2^*}}\right) \cdot \frac{1}{2}\, d\lambda
\end{aligned}$$

with

$$\lambda_1^* = \sqrt{2 - \sqrt{4 - \lambda'^2}} \tag{41}$$

$$\lambda_2^* = \sqrt{2 + \sqrt{4 - \lambda'^2}}. \tag{42}$$

The result is, continuing to indicate in this subsection the chord of interest with λ',

$$f(\lambda'\,|\,\mathcal{M}_5) = \frac{\sqrt{2 + \sqrt{4 - \lambda'^2}} + \sqrt{2 - \sqrt{4 - \lambda'^2}}}{4\sqrt{4 - \lambda'^2}}. \tag{43}$$

$\overline{GF}$. It follows then

$$l' = 2 \times l\sqrt{l^2/2R \cdot (2R - l^2/2R)}.$$

[9]It is remarkable that $\sqrt{2 + \sqrt{3}} - \sqrt{2 - \sqrt{3}} = \sqrt{2}$, an identity used to rewrite Eq. (40) in a more compact form.

17

The cumulative distribution can be obtained making the usual integral, although the integrand is in this case particularly nasty.[10] In reality the cumulative distribution can be calculated extending the reasoning followed above to calculate the probability that the chord is smaller than the side of the triangle and than the radius. In fact, remembering the transformation from λ to λ' plotted in Fig. 9 and using capital letters to distinguish the variables from their values, we have

$$P(\Lambda' \leq \lambda') \;=\; P\left(\Lambda \leq \sqrt{2 - \sqrt{4 - \lambda'^2}}\right) + P\left(\sqrt{2 + \sqrt{4 - \lambda'^2}} \leq \Lambda \leq 2\right) \quad (44)$$

$$=\; \frac{\sqrt{2 - \sqrt{4 - \lambda'^2}}}{2} + \frac{2 - \sqrt{2 + \sqrt{4 - \lambda'^2}}}{2}\,, \quad (45)$$

from which it follows

$$F(\lambda' \,|\, \mathcal{M}_5) \;=\; \frac{1}{2}\left(2 + \sqrt{2 - \sqrt{4 - \lambda'^2}} - \sqrt{2 + \sqrt{4 - \lambda'^2}}\right)\,, \quad (46)$$

that, derivated, reproduces Eq. (43). Pdf and cumulative functions are shown in Fig. 10. The two probabilities of interest are then

$$P(\lambda' \leq \sqrt{3} \,|\, \mathcal{M}_5) \;=\; \frac{\mathbf{3} - \sqrt{\mathbf{3}}}{\mathbf{2}} \approx 0.634 \quad (47)$$

$$P(\lambda' \leq 1 \,|\, \mathcal{M}_5) \;=\; 1 - \frac{1}{\sqrt{2}} \approx 0.293\,. \quad (48)$$

Results based on simulations are shown in Figs. 19-23.

3.3   Summary of the results obtained by the various methods

The results obtained by the 'geometric' methods $\mathcal{M}_1$-$\mathcal{M}_5$ are summarized in Tab. 1 and in Fig. 11, together with the more physical one $\mathcal{M}_6$, that will be discussed in section 6. In the table we have also added, for completeness, the expected values and the standard deviations of the probability distributions. More details, obtained from simulations, are shown in Figs. 19-23. The dashed lines of Fig. 11 represent instead the arithmetic averages of the other functions, whose motivation will be clear in Appendix 3.

---

[10]Wolfram Mathematica provides the following result:

$$F(\lambda' \,|\, \mathcal{M}_5) \;=\; \frac{1}{2\lambda'}\left(2\lambda' - 2\sqrt{2 - \sqrt{4 - \lambda'^2}} - \sqrt{(4 - \lambda'^2) \cdot (2 - \sqrt{4 - \lambda'^2})}\right.$$

$$\left. + 2\sqrt{2 + \sqrt{4 - \lambda'^2}} - \sqrt{(4 - \lambda'^2) \cdot (2 + \sqrt{4 - \lambda'^2})}\right)\,,$$

that, although apparently quite different from Eq. (46), can be checked to be numerically equivalent to it.
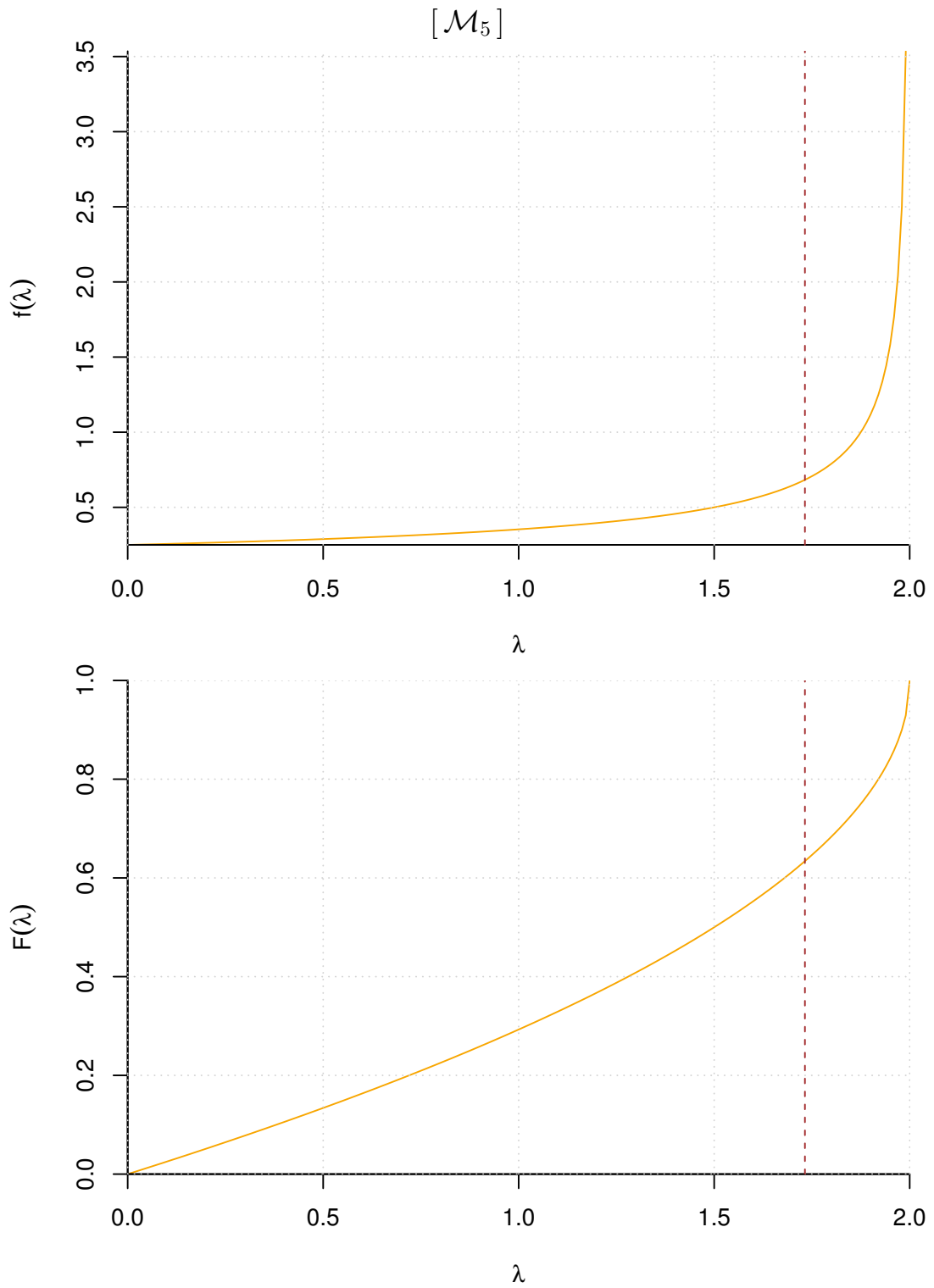
Figure 10: Probability distribution of $\lambda = l/R$ of the chords generated with Method 5. The dashed vertical line indicates $\lambda = \sqrt{3}$.

Model	$f(\lambda)$	$F(\lambda)$	$E(\lambda)$	$\sigma(\lambda)$	$\mathbf{P(\lambda \leq \sqrt{3})}$	$P(\lambda \leq 1)$
$\mathcal{M}_1$	$\dfrac{1}{\pi\sqrt{1-(\lambda/2)^2}}$	$\dfrac{2}{\pi}\cdot\arcsin\dfrac{\lambda}{2}$	$4/\pi$ (≈ 1.27)	$\sqrt{2-16/\pi^2}$ (≈ 0.62)	$\mathbf{2/3}$ (≈ 0.67)	$1/3$ (≈ 0.33)
$\mathcal{M}_2$	$\dfrac{\lambda}{4\sqrt{1-(\lambda/2)^2}}$	$1-\sqrt{1-(\lambda/2)^2}$	$\pi/2$ (≈ 1.57)	$\sqrt{8/3-\pi^2/4}$ (≈ 0.45)	$\mathbf{1/2}$ ($=0.5$)	$1-\sqrt3/2$ (≈ 0.13)
$\mathcal{M}_3$	$\lambda/2$	$\lambda^2/4$	$4/3$ (≈ 1.33)	$\sqrt2/3$ (≈ 0.47)	$\mathbf{3/4}$ ($=0.75$)	$1/4$ ($=0.25$)
$\mathcal{M}_4$	$1/2$	$\lambda/2$	1	$1/\sqrt3$ (≈ 0.58)	$\mathbf{\sqrt3/2}$ (≈ 0.87)	$1/2$ ($=0.5$)
$\mathcal{M}_5$	[*]	[**]	$4/3$ (≈ 1.33)	$4/(3\sqrt5)$ (≈ 0.60)	$\mathbf{(3-\sqrt3)/2}$ (≈ 0.63)	$1-1/\sqrt2$ (≈ 0.29)
$\mathcal{M}_6$	$\dfrac{2}{\pi}\dfrac{(\lambda/2)^2}{\sqrt{1-(\lambda/2)^2}}$	[***]	$16/3\pi$ (≈ 1.70)	$\sqrt{3-256/9\pi^2}$ (≈ 0.34)	$\mathbf{2/3-\sqrt3/2\pi}$ (≈ 0.39)	$1/3-\sqrt3/2\pi$ (≈ 0.058)

[*] $\quad f(\lambda\,|\,\mathcal{M}_5) = \left(\sqrt{2+\sqrt{4-\lambda^2}}+\sqrt{2-\sqrt{4-\lambda^2}}\,\right)/(4\sqrt{4-\lambda^2})$

[**] $\quad F(\lambda\,|\,\mathcal{M}_5) = \frac{1}{2}\left(2+\sqrt{2-\sqrt{4-\lambda^2}}-\sqrt{2+\sqrt{4-\lambda^2}}\,\right)$

[***] $F(\lambda\,|\,\mathcal{M}_6) = \frac{2}{\pi}\arcsin(\lambda/2)-\frac{\lambda}{\pi}\sqrt{1-(\lambda/2)^2}$

Table 1: Summary of the results obtained by the different 'random' drawing models of the chords
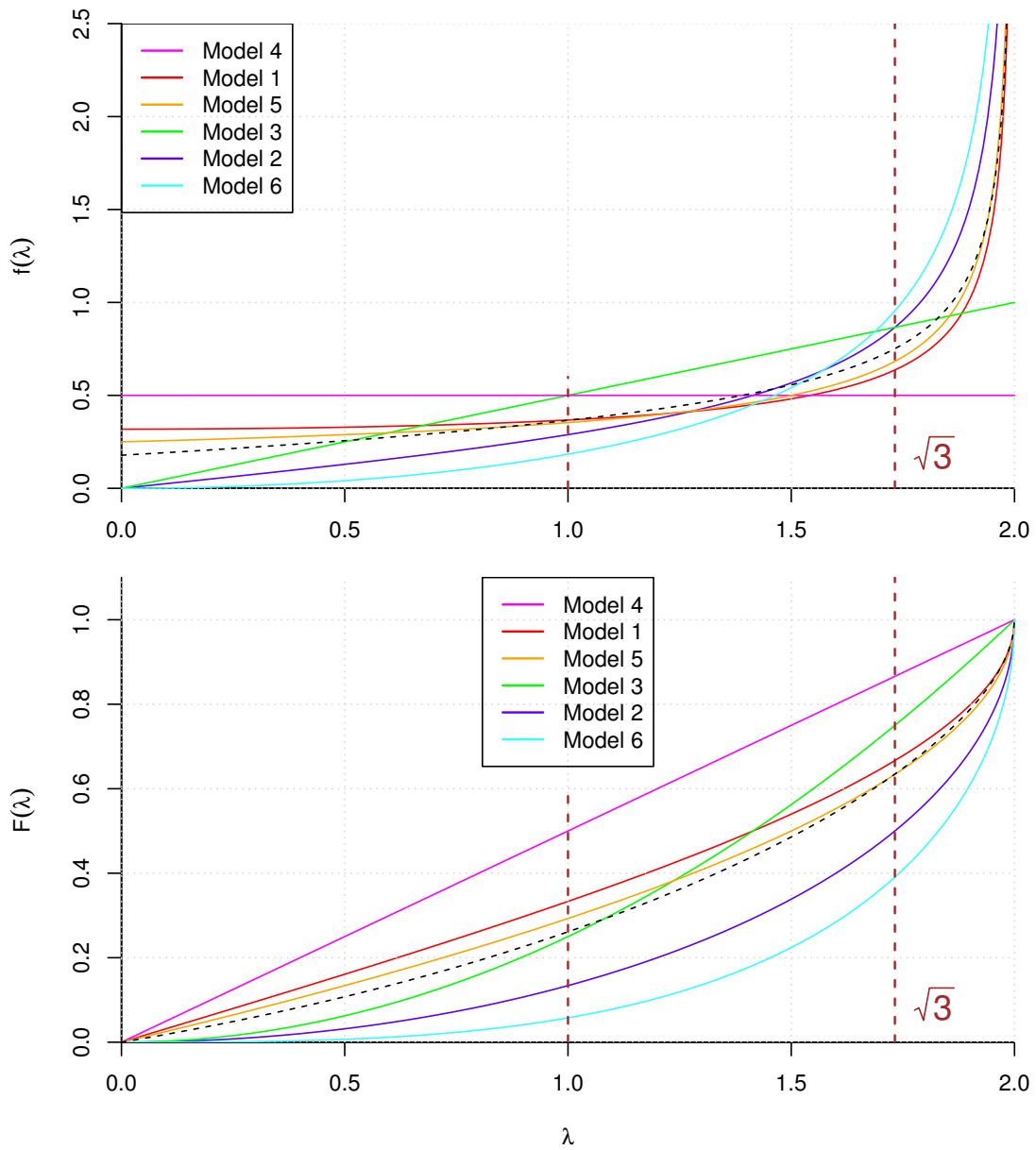
Figure 11: Probability density functions and cumulative function of the lengths of the chords in units of the radius of the circle for the extraction models considered in the text (the order of the models in the legend corresponds to the decreasing order of $F(1 \,|\, \mathcal{M}_i)$ and $f(0.1 \,|\, \mathcal{M}_i)$). The dashed lines correspond to the 'averages', on which we shall come back in Appendix C.

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

In reality, if we are just interested in writing a computer program which generates the length of the chords 'at random', using one of the five 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.

4.1 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 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

$$y \;=\; g(x) = F_X(x)\,, \tag{49}$$

written in 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 the pdf of Y can be evaluated as

$$f_Y(y) \;=\; \int_{-\infty}^{+\infty} \delta\left(y - F_X(x)\right) \cdot f_X(x)\, dx\,. \tag{50}$$

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

$$f_Y(y) \;=\; \int_{-\infty}^{+\infty} \frac{\delta(x - x^*)}{F_X'(x^*)} \cdot f_X(x)\, dx = \frac{f_X(x^*)}{f_X(x^*)} = 1\,. \tag{51}$$

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

$$F_Y(y) \equiv P(Y \le y) \;=\; P(Y \le F_X(x)) = P(X \le F_X^{-1}(y)) \equiv F_X\left(F_X^{-1}(y)\right) = y$$

$$f_Y(y) \equiv \frac{d}{dy} F_Y(y) \;=\; \frac{d}{dy} y = 1\,.$$

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 be inverted*, to make a (pseudo) random number generator to produce numbers such that our confidence on the occurrence of $X = x$ is proportional to $f(x)$:

$$x \;=\; F_X^{-1}(u)\,, \tag{52}$$

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.)

4.2 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 (52) gets then implemented as follows.

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

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

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

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

$\mathcal{M}_5$: $\lambda = 2\sqrt{1 - 2u^2 + u^4}$.

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,
         2*sqrt(1-2*u^2+u^4))
}
```

Issuing then the following instruction from an R console ('>' stands for the prompt) you can for example get an histogram similar to the left top one of Fig. 20:
```
> lambda=rlchords(100000,1); hist(lambda, nc=200, col='red')
```
Or you can calculate sample mean and standard deviation, and fraction of occurrences with λ smaller than $\sqrt{3}$ with these very simple commands
```
> mean(lambda)
```
```
> sd(lambda)
```
```
> length(lambda[lambda<sqrt(3)])/length(lambda)
```
(You will get sample mean and standard deviation very close to the mean and standard deviation of the distribution *because it is very unlikely* to have something very different, as famous a theorem of probability theory makes us feel confident.)

5   What should one expect from a computer drawing program?

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 <u>draw</u> the chord, by hand of by a computer program.[11] 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 circle. Indeed – I have to confess with some shame – the first time I was playing with the Bertrand 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 <u>not</u> 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. In that way it was possible to reuse somehow the 'blocks' (the graphical programming elements shown e.g. in Fig. 12) 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. (My suggestion to students is that the for everyday use the 'fastest' algorithm is the one that they write more rapidly and understand better – unless you need it for special purposes, it is a waste of time to spend several hours of your life to write a piece of program that provides the result in microseconds instead then in tens, hundreds or even thousands of milliseconds).

---

[11]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
```

5.1 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, that should be set in a previous command).

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

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

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

In this case we start extracting $x_1 = x_2$ between -1 and 1 and evaluate the corresponding ordinates on the circumference, i.e.

```
> p1 <- p2 <- runif(n, -1, 1)
> p1 <- cbind(p1, sqrt(1-p1^2))    # [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 can reuse exactly the last three lines of code of the previous method, i.e. starting from the line tagged by "`# [1]`". The resulting App Inventor blocks are shown in Fig. 12.

5.3 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 4.1.

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

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

```
> 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, tagged by "`# [2]`". The implementation of this variation in App Inventor is shown in Fig. 13.
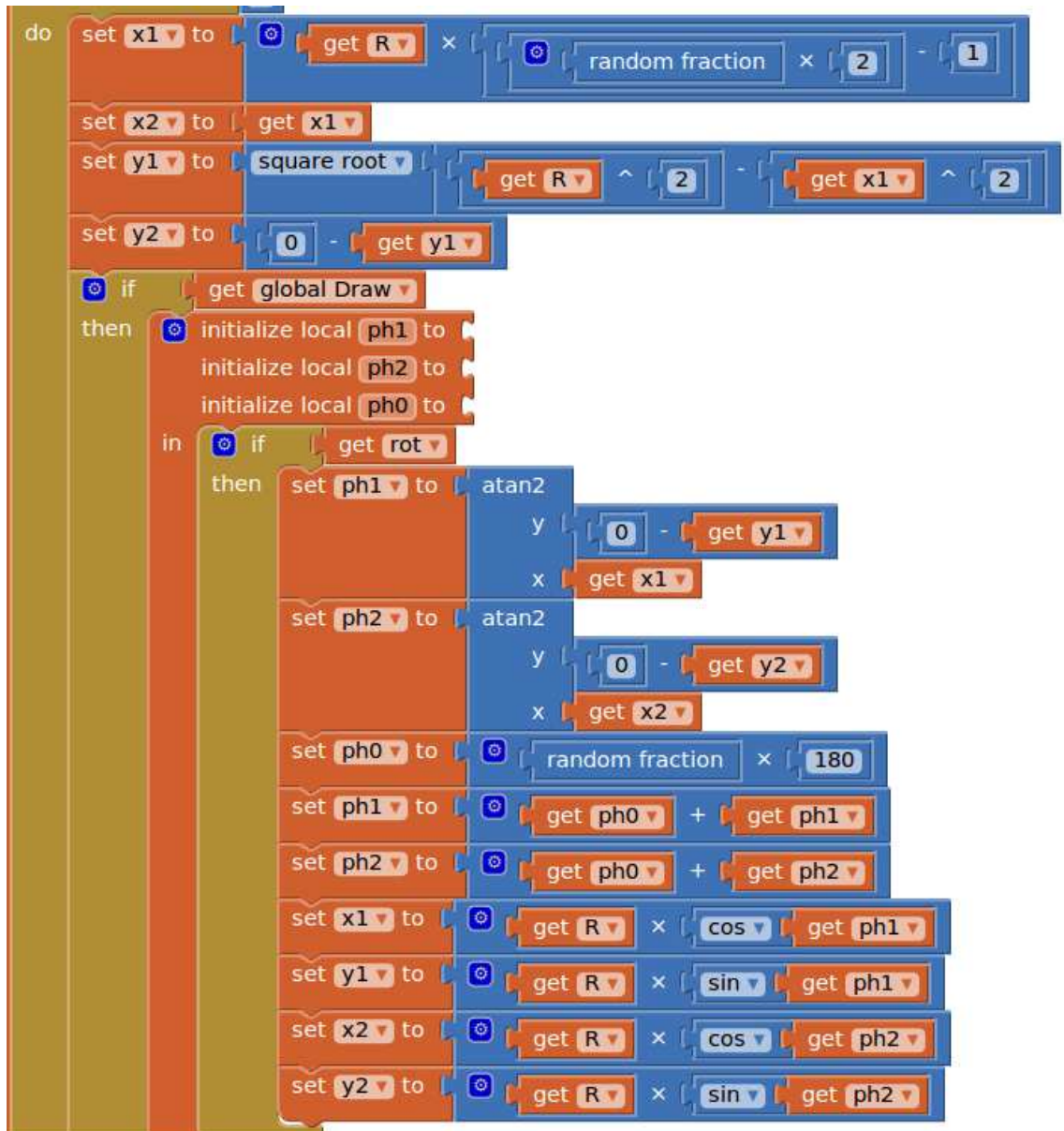
Figure 12: App Inventor blocks for the core of Method 2. (The minus sign of the y argument of atan2 is due to the fact that in App Inventor the y coordinate inside a 'canvas' is upside down, being the origin in the top left corner, while the 'heading' angle follows the usual convention. Note also how, contrary to other scientific libraries, the trigonometric functions use degrees.)
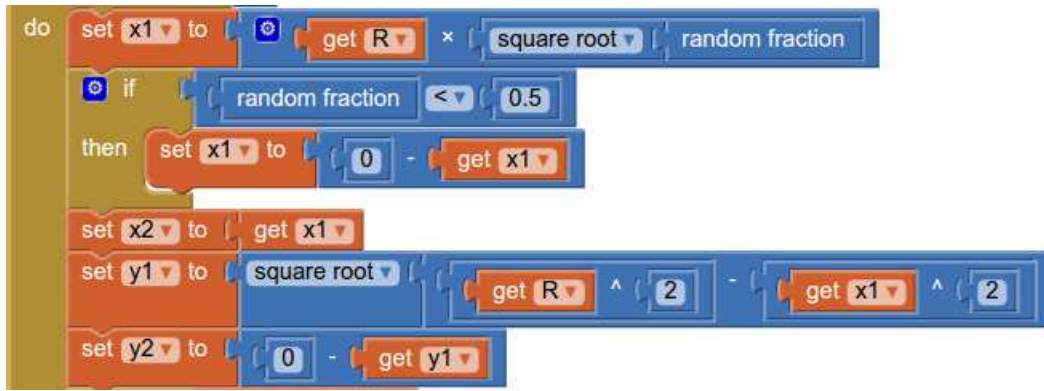
Figure 13: App Inventor blocks to turn Method 2 into Method 3 (see text).

5.4 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. In fact the first endpoint is $(-R, 0)$, while the second can be easily found using elementary geometry. With the help of Fig. 14 we can recognize a useful rectangular triangle and then make use of a famous theorem of geometry. The projection p is then $l^2/2R$ and then the abscissa x_2 of the intersection is equal to $x_2 = p - R = l^2/2R - R$. The corresponding ordinate is therefore $y_2 = \sqrt{R^2 - x_2^2}$. Then we can rotate 'at random' the points as done above, when we described Method 2. These are the two lines of R code that replace the <u>first</u> line of code above

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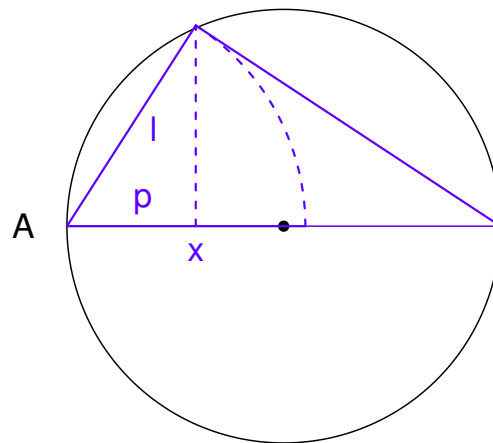

Figure 14: A sketch which show how to calculate the abscissa of the the interception (x) by simple geometry (see text).

and from the second line, tagged by "`# [2]`", it will be all the same.

5.5 Model 5 ($\mathcal{M}_5$)

The difference with the previous model is that now the abscissa of the two points (before rotation!) are the same, while the ordinates are one the opposite opposite of the other, as in Model 2. We have then

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

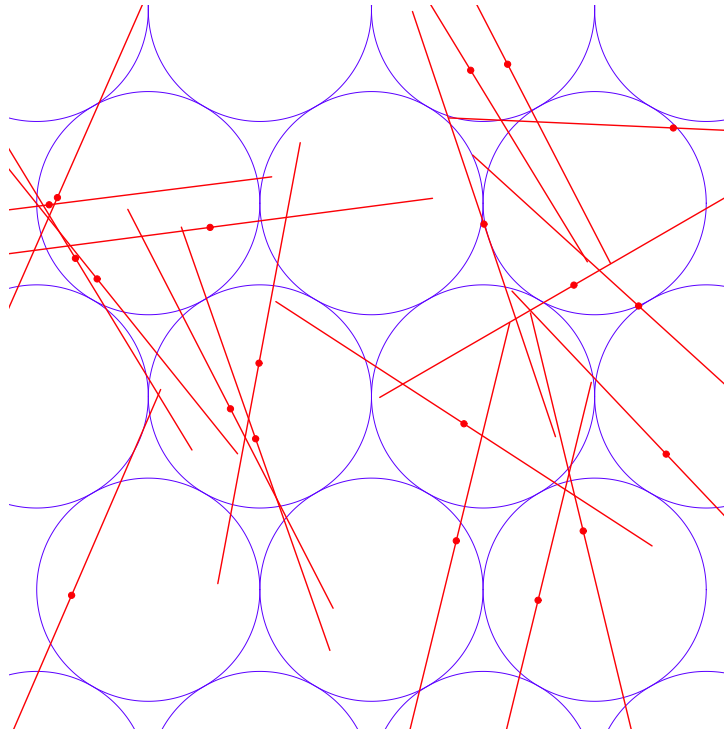and then we continue from the second line of the code of Model 2, tagged by "`# [2]`".

Figure 15: Mikado sticks thrown at random on a pattern of circles.

6 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. 15: a pattern of circles [12] on a table, or on the floor, on

---

[12]In Fig. 15 all circles have the same size, but this is not a necessary requirement, as it will be clear in a while.
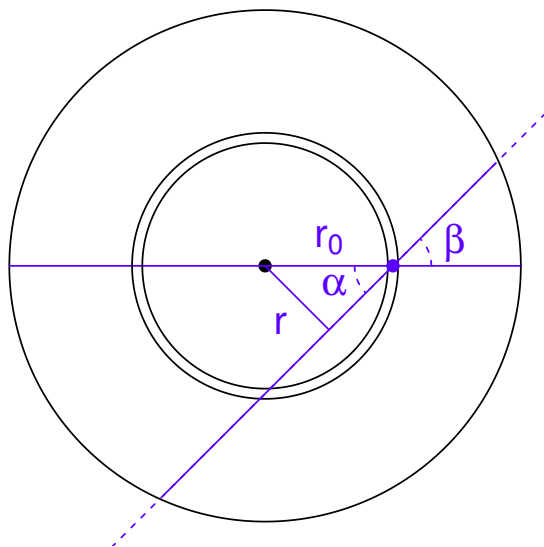
Figure 16: Sketch of the mikado experiment to show how to evaluate the distance of the chord from the center from the position of the center of the stick and its orientation (see text).

which we throw mikado sticks, or toothpicks, or needles or something similar. The only, obvious, conditions is on the minimum length of the sticks, that has to be at least twice the maximum diameter of the circles.[13]

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 intersection of the stick with the circumference of that circle. As we can see from Fig. 15, 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 a different 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 $f(\rho_0 \,|\, \mathcal{M}_6) = 2\rho_0$. 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

_____

[13]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".

our starting conditions [14]

$$\rho = \rho_0 \sin \alpha \tag{53}$$

$$f(\rho_0 \,|\, \mathcal{M}_6) = 2\rho_0 \qquad (0 \le \rho_0 \le 1) \tag{54}$$

$$f(\alpha \,|\, \mathcal{M}_6) = \frac{1}{\pi/2} = \frac{2}{\pi} \qquad (0 \le \alpha \le \frac{\pi}{2}). \tag{55}$$

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

$$f(\rho \,|\, \mathcal{M}_6, \rho_0) = \int_0^{\pi/2} \delta(\rho - \rho_0 \sin \alpha) \cdot f(\alpha) \, d\alpha \tag{56}$$

$$= \int_0^{\pi/2} \frac{\delta(\alpha - \alpha^*)}{\rho_0 \cos \alpha^*} \cdot \frac{2}{\pi} \, d\alpha, \tag{57}$$

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

$$f(\rho \,|\, \mathcal{M}_6, \rho_0) = \frac{2}{\pi \, \rho_0 \sqrt{1 - (\rho/\rho_0)^2}} \tag{58}$$

$$= \frac{2}{\pi \, \sqrt{\rho_0^2 - \rho^2}}. \tag{59}$$

Having $f(\rho \,|\, \mathcal{M}_6, \rho_0)$ and $f(\rho_0 \,|\, \mathcal{M}_6)$ we can then evaluate $f(\rho \,|\, \mathcal{M}_6)$ as

$$f(\rho \,|\, \mathcal{M}_6) = \int_\rho^1 f(\rho \,|\, \mathcal{M}_6, \rho_0) \cdot f(\rho_0 \,|\, \mathcal{M}_6) \, d\rho_0, \tag{60}$$

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

$$f(\rho \,|\, \mathcal{M}_6) = \int_\rho^1 \frac{2}{\pi \, \sqrt{\rho_0 - \rho^2}} \cdot 2\rho_0 \, d\rho_0 \tag{61}$$

$$= \int_\rho^1 \frac{4 \, \rho_0}{\pi \, \sqrt{\rho_0^2 - \rho^2}} \, d\rho_0 \tag{62}$$

$$= \frac{4}{\pi} \sqrt{1 - \rho^2}. \tag{63}$$

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

$$f(\lambda \,|\, \mathcal{M}_6) = \int_0^1 \delta\left(\lambda - 2\sqrt{1 - \rho^2}\right) \cdot \frac{4}{\pi} \sqrt{1 - \rho^2} \, d\rho \tag{64}$$

---

[14] An alternative way would be to use the angle β defined in Fig. 16, ranging between 0 to π, with $f(\beta \,|\, \mathcal{M}_6) = 1/\pi$. 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 (53) would then be replaced by $\rho = \rho_0 \sin \beta$ and the integral (56) 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.

$$= \int_0^1 \frac{\delta\left(\rho - \rho^*\right)}{2\rho^2/\sqrt{1-\rho^2}\Big|_{\rho=\rho^*}} \cdot \frac{4}{\pi}\sqrt{1-\rho^2}\, d\rho \tag{65}$$

$$= \frac{4/\pi\sqrt{1-\rho^{*2}}}{2\rho^*/\sqrt{1-\rho^{*2}}} = \frac{2}{\pi}\frac{1-\rho^{*2}}{\rho^*}, \tag{66}$$

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

$$f(\lambda\,|\,\mathcal{M}_6) \;=\; \frac{2}{\pi}\frac{(\lambda/2)^2}{\sqrt{1-(\lambda/2)^2}} \tag{67}$$

$$F(\lambda\,|\,\mathcal{M}_6) \;=\; \frac{2}{\pi}\arcsin\left(\lambda/2\right) - \frac{\lambda}{\pi}\sqrt{1-(\lambda/2)^2}\,, \tag{68}$$

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

$$\mathrm{E}(\lambda\,|\,\mathcal{M}_6) \;=\; \frac{16}{3\,\pi} \approx 1.70 \tag{69}$$

$$\sigma(\lambda\,|\,\mathcal{M}_6) \;=\; \sqrt{3 - \frac{256}{9\,\pi^2}} \approx 0.34 \tag{70}$$

$$P(\lambda \leq \sqrt{3}\,|\,\mathcal{M}_6) \;=\; \frac{2}{3} - \frac{\sqrt{3}}{2\,\pi} \approx 0.39 \tag{71}$$

$$P(\lambda \leq 1\,|\,\mathcal{M}_6) \;=\; \frac{1}{3} - \frac{\sqrt{3}}{2\,\pi} \approx 0.058\,. \tag{72}$$

6.1 Remarks on simulation

We have seen in section 4 how to write chord length generators for the different methods without having to go through the steps of drawing the chords, while we have learned in section 5 how to reuse the code to draw the chords. In this case the situation seems a bit more complicate, but there are fortunately some ways out.

6.1.1 Random chord length generator

The generators described in subsection 4.2 were based on the trick of inverting the cumulative distribution. However Eq. (68) is not easily invertible and we cannot use that trick.[15] We could then go one step behind, and try to generate the values of ρ, normalized distance of the chords from the center, and then calculate, from each ρ, the corresponding λ. But also in this case the cumulative distribution of interests is not invertible, being equal to

$$F(\rho,\,|\,\mathcal{M}_6) \;=\; \frac{2}{\pi}\left(\rho\sqrt{1-\rho^2} + \arcsin\rho\right). \tag{73}$$

We could make use of the *hit/miss* method, simply introduced with the help of Fig. 18. The left plot shows (black curve) the pdf $f(\rho)$ inside a rectangle defined by the range

---

[15]One might think to use the *hit/miss* method, described in the text in the following lines. But this is also not possible to be used since $f(\lambda)$ is divergent for $\lambda \to 2$, although its integral is finite.
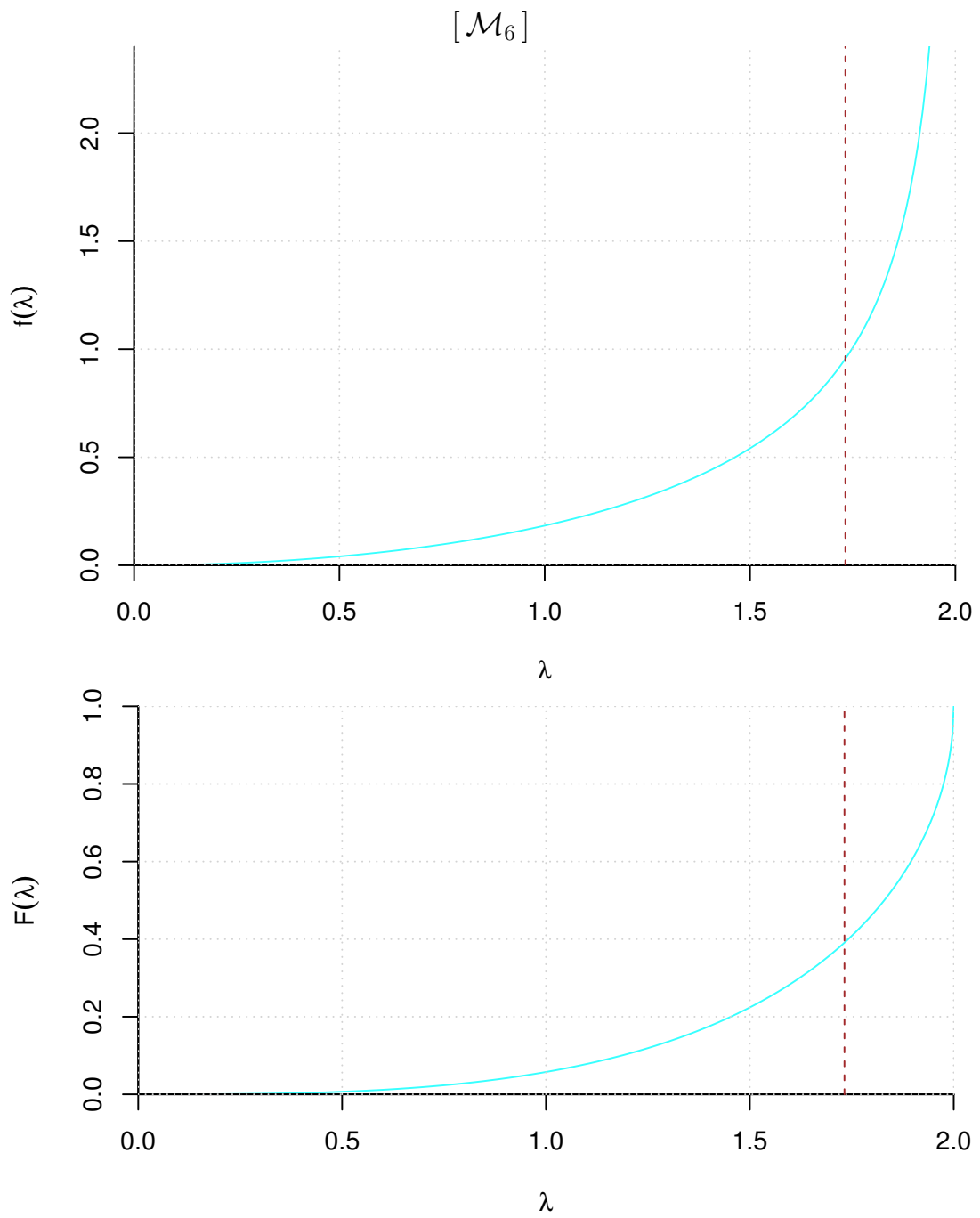
Figure 17: Probability distribution of $\lambda = l/R$ of the chords generated with Method 6 ('mikado'). The dashed vertical line indicates $\lambda = \sqrt{3}$.
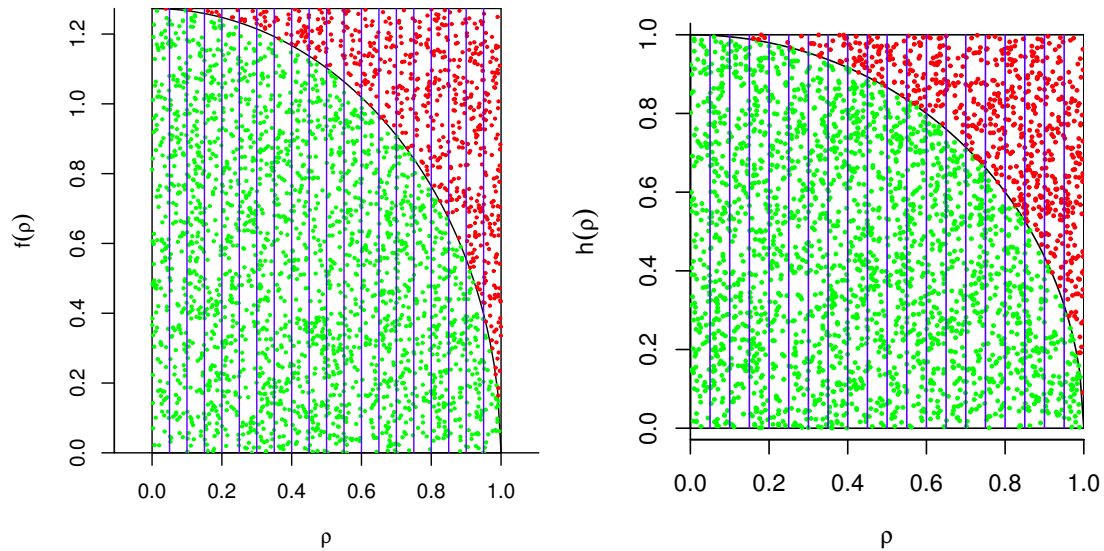
Figure 18: Extracting ρ according to its pdf using the hit/miss Monte Carlo technique. (But we shall make direct extractions inside the quarter of circle! See text.)

of ρ and the range of its pdf. We then throw 'at random' points uniformly inside the rectangle and mark in green those which fall between $f(\rho)$ and the abscissa, in red those which fall 'outside'. We draw then some vertical lines to identify 'slides' in ρ. We see that the number of the green point in each slice is proportional to $f(\rho)$ calculated in the middle of the slide. We can imagine then the slides very thin, with the usual procedures done in calculus, to convince ourselves that the probability that a green point falls inside a slide is proportional to $f(\rho)$. This is a well known technique to make extractions according to a given pdf, at the expenses of some inefficiency, which in our case is tolerable (just the fraction of red points over the total). It becomes instead not tolerable if the pdf is very peaked somewhere and assumes very small values in the rest of the range of the variable. Or it becomes impossible to be used when the pdf diverges and the rectangle become infinite high, as it would be with $f(\lambda \,|\, \mathcal{M}_6)$.

To conclude, in our case the hit/miss technique would be appropriate, but we can do even better, if we observe that the factor $4/\pi$ in Eq. (63) is irrelevant for the reasoning. If we drop it, we are left with $\sqrt{1-\rho^2}$, which describe a circle in the first quadrant, as shown in the right plot of Fig. 18, with the ordinate indicated now by $h(\rho)$, being a generic 'height' and not a pdf. But we already know how to extract points uniformly inside a circle without having to throw points in the square circumscribed! What we need is just to limit the extraction in a quarter of the circle, and this can be easily achieved limiting the polar angle between 0 and $\pi/2$. Here is directly the R command to generate a single value of ρ, followed by the corresponding value of λ

```
> rho=sqrt(runif(1))*cos(runif(1,0,pi/2))
> lambda <- 2 * sqrt(2 - rho^2)
```

and finally in a single step, with n values:

33

```
> lambda <- 2 * sqrt(2 - runif(n)*cos( runif(n,0,pi/2)^2 )
```
Here is then the new version of our generator with all six models implemented:

```
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,
         2*sqrt(1-2*u^2+u^4),
         2*sqrt(1-u*cos(runif(n, 0, pi/2))^2))
}
```

6.1.2   Random chord drawer

Sampling ρ is the key to simulate chords inside the circle selected by the mikado reference point, allowing us to reuse the code written for other methods. We can then easily extend what we have done in section 5. All what we have to do is to replace in the code for Model 3 (subsection 5.3) the line
```
> rho <- sqrt(runif(n))
```
by
```
> rho <- sqrt(runif(n, 0, 1))*cos(runif(n, 0, pi/2))
```
and the game is over. Results from simulations are shown in Figs. 19-23.

7   Conclusions

Arguing, in abstract terms, about *the* solution of the Bertrand problem – not a paradox! – is as scientific as the Byzantine debates about the sex of angels. Nevertheless, the question can still have a sense if framed in a practical contest, asking people to draw, by hand or with the help of computer graphics, <u>a</u> chord and betting on the result.

Several random drawing methods have been analyzed in great detail, in order to show several issues related to the evaluation of the probability distributions of functions of variables whose pdf was assumed and to the drawing simulation.

Of the six methods analyzed, five can be classified as 'geometrical', two of which are the 'classical' solutions. There was no intention to invent abstruse methods and indeed the two "with ruler and compass" algorithms seem to me even more 'natural' (or at least more suited to a class of people) that the more famous, 'classical' Method 3.

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

<p align="center">*— Il faut cultiver notre jardin —*</p>

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 Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria (2014). http://www.R-project.org/

[7] D. Wolber et al., *App Inventor 2*, O'Reilly Media, 2 edition (2014). http://ai2.appinventor.mit.edu/

[8] G. D'Agostini, *"Teaching statistics in the physics curriculum. Unifying and clarifying role of subjective probability"*, Am. J. Phys. **67** (1999) 1260 [physics/9908014].
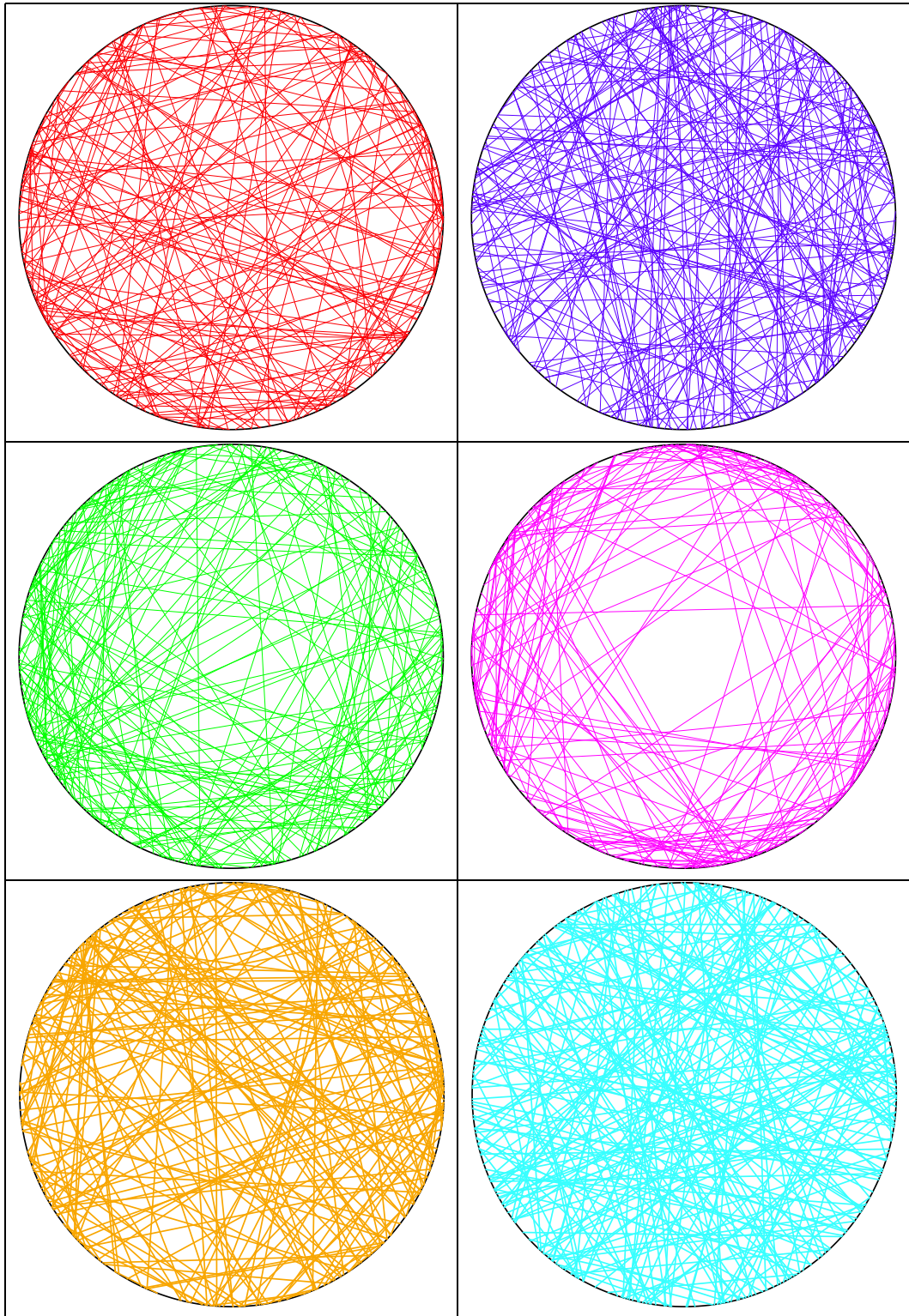
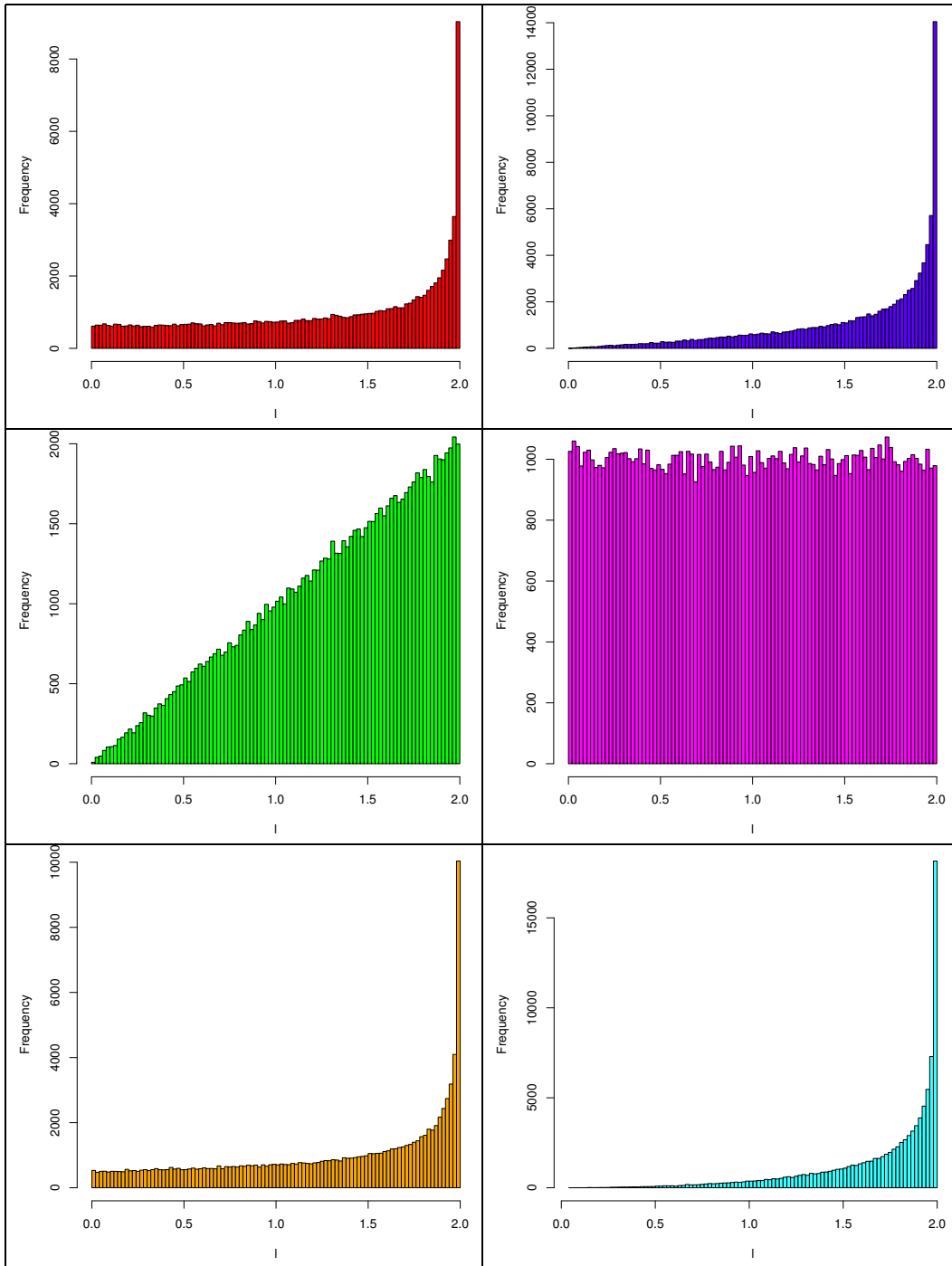Figure 19: Samples of chords generated the various methods.

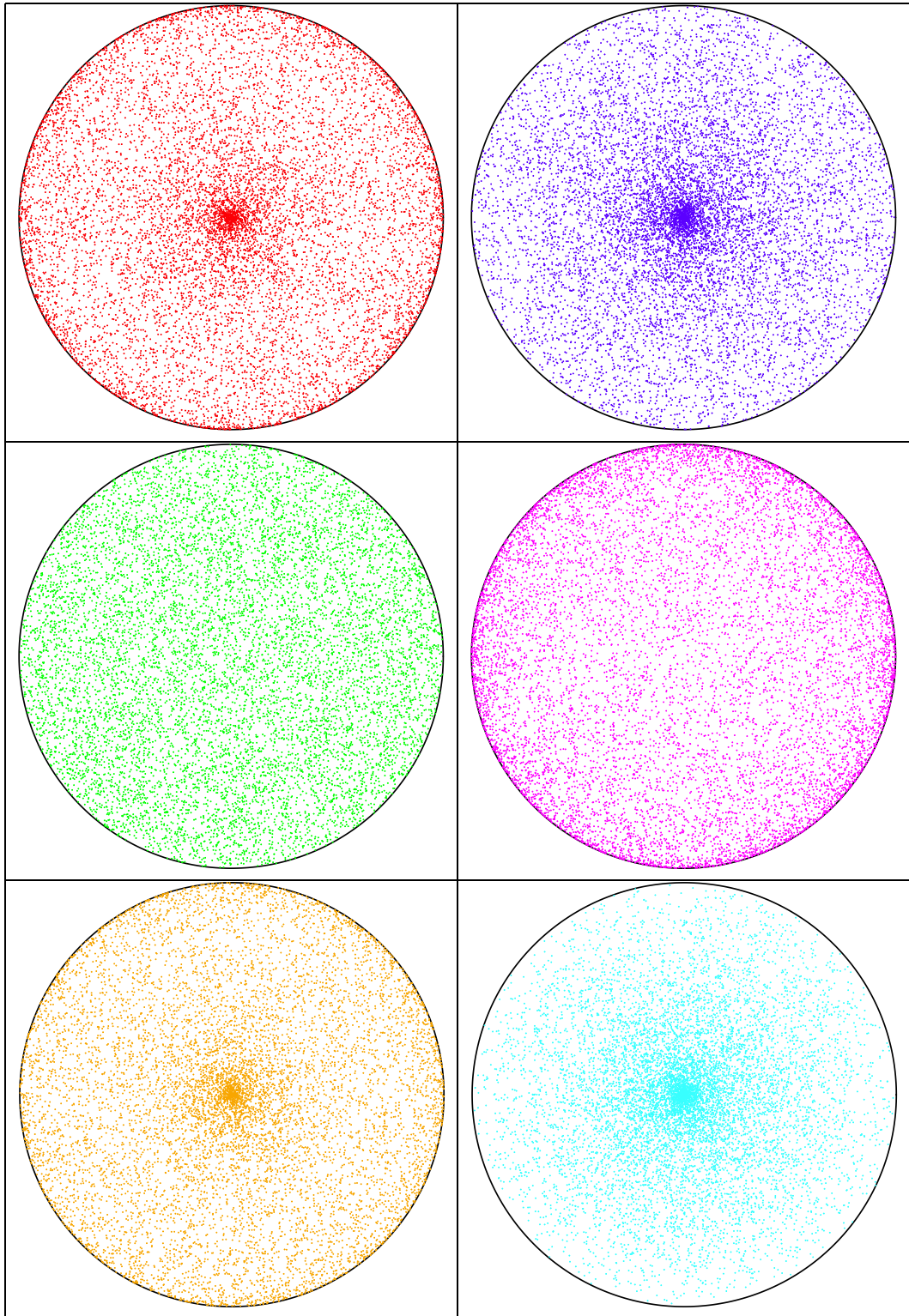Figure 20: Distribution of the length of the chords in sample produced with the various methods.

Figure 21: Sample of centers of the chords generated with the various methods
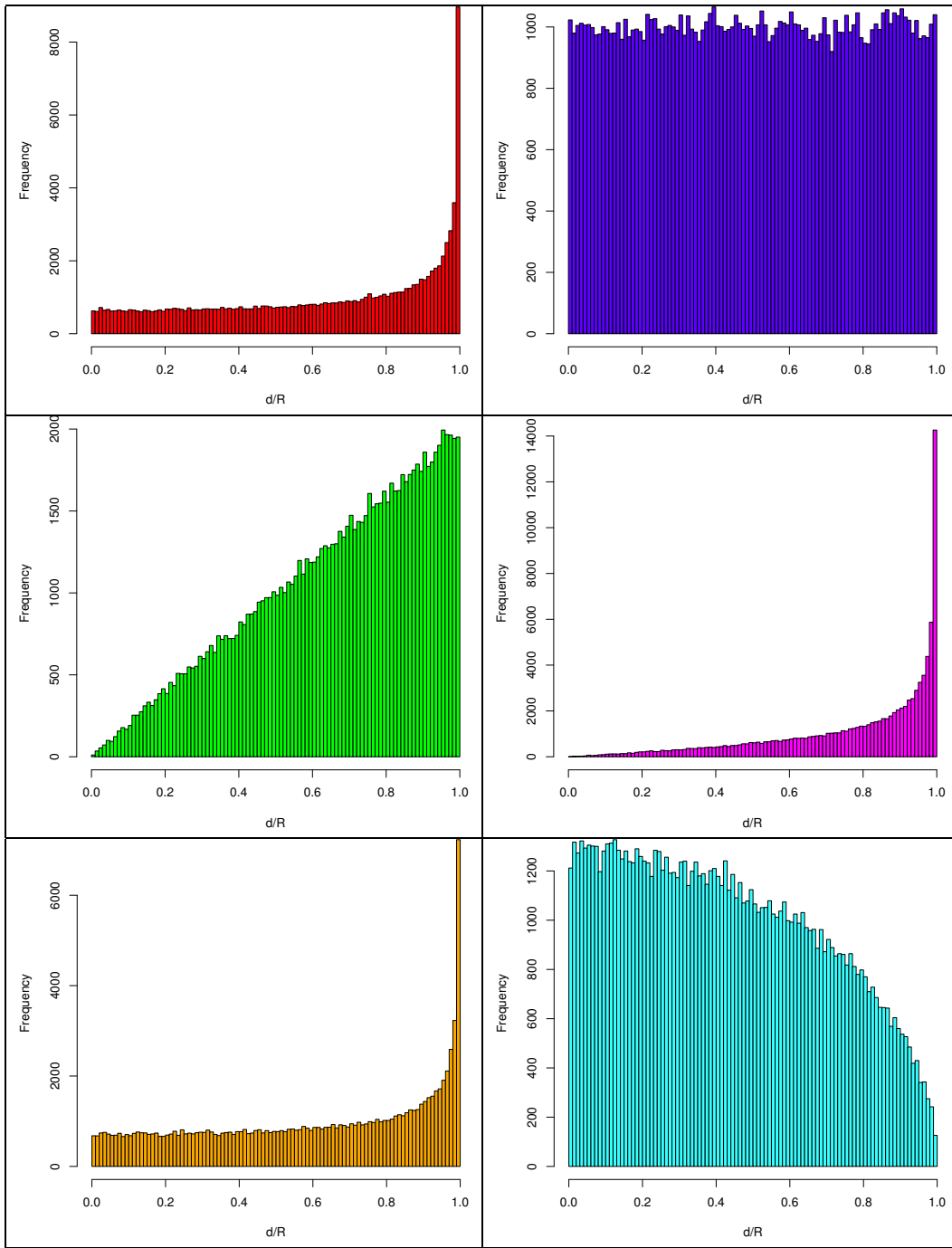
Figure 22: Distribution of the distance of the chords from center of the circle in samples produced with the various methods.
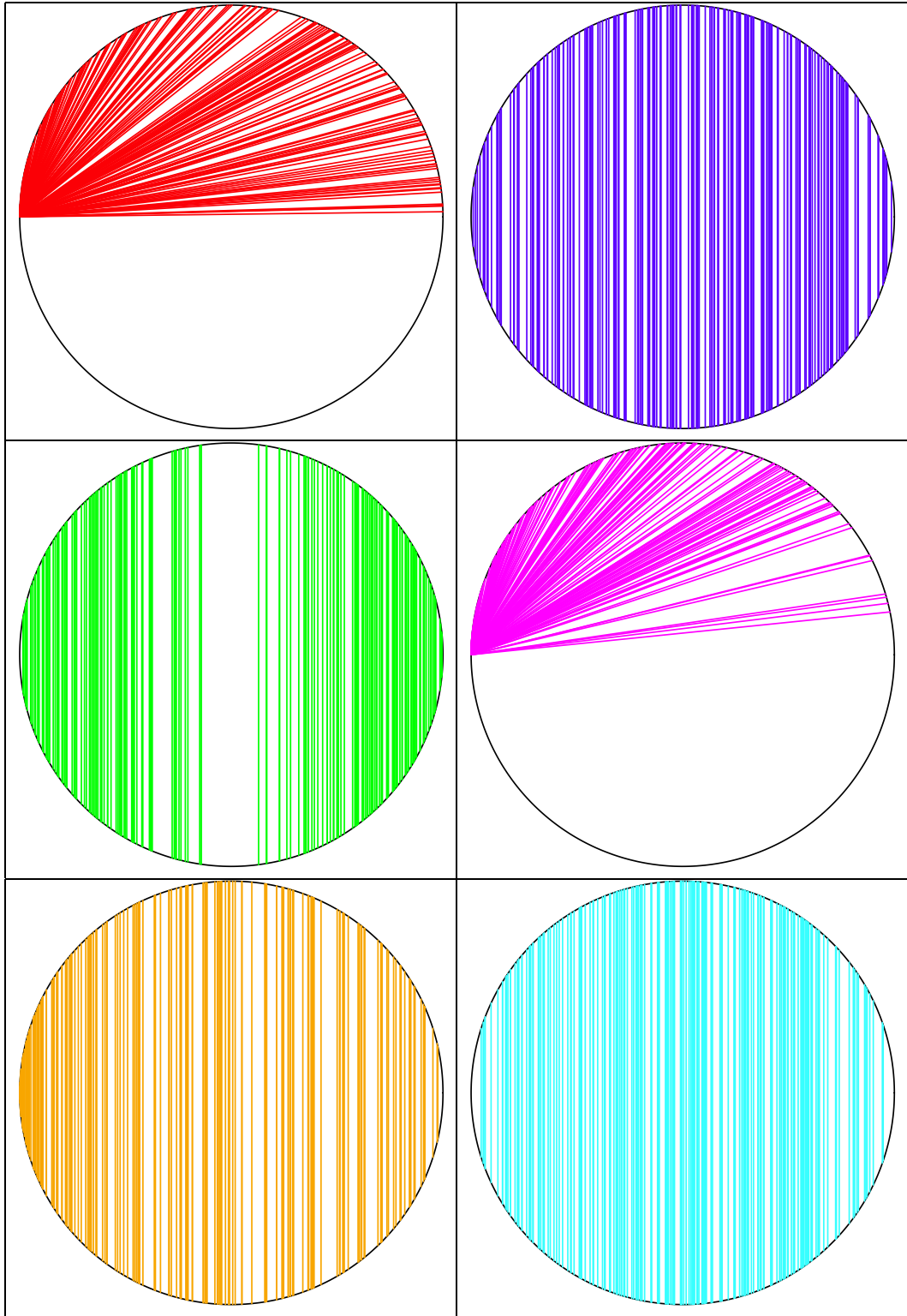
Figure 23: Samples of chords generated the various methods with respect to a preferred axis ("without rotation" – see text).

Appendix A: Distributions of chord distances from the center of the circle

Out of the six methods analyzed above, one ($\mathcal{M}_2$) is defined in terms of the distance between chord and center of the circle, while in two others ($\mathcal{M}_3$ and $\mathcal{M}_6$) we have learned that that distance is a useful quantity to simplify the solution of the problem. The three pdf's of the normalized distances in these three cases are

$$f(\rho\,|\,\mathcal{M}_2) \;=\; 1 \tag{74}$$

$$f(\rho\,|\,\mathcal{M}_3) \;=\; 2\,\rho \tag{75}$$

$$f(\rho\,|\,\mathcal{M}_6) \;=\; \frac{4}{\pi}\,\sqrt{1-\rho^2}\,, \tag{76}$$

while we have shown in Fig. 22 the results of the simulation for all six methods. For completeness, let us do the exercise to calculate the pdf $f(\rho)$ also for the other three methods. This is in fact easy, at least in principle, because there a simple geometric relation between the length of a chord and its distance from the center, since the chord and the radii to its endpoints identify an equilateral triangle. The only technical problem is that of getting a closed expression for the pdf. Let us remind here the relations between length λ of the chord and distance ρ from the center, both in units of the radius.

$$\rho \;=\; \sqrt{1-(\lambda/2)^2} \tag{77}$$

$$\lambda \;=\; 2\,\sqrt{1-\rho^2}\,, \tag{78}$$

which we shall use in the following subsections. In fact, using our transformation rule with the Dirac delta, $f(\rho)$ is given by

$$
\begin{aligned}
f(\rho\,|\,\mathcal{M}_i) \;&=\; \int_0^2 \delta\!\left(\rho-\sqrt{1-(\lambda/2)^2}\right)\cdot f(\lambda\,|\,\mathcal{M}_i)\,d\lambda \tag{79}\\[2mm]
&=\; \int_0^2 \delta\!\left(\rho-\sqrt{1-(\lambda/2)^2}\right)\frac{\delta(\lambda-\lambda^*)}{\left|\frac{d}{d\lambda}\sqrt{1-(\lambda/2)^2}\right|_{\lambda=\lambda^*}}\cdot f(\lambda\,|\,\mathcal{M}_i)\,d\lambda\\[2mm]
&=\; \frac{4\,\sqrt{1-(\lambda^*/2)^2}}{\lambda^*}\cdot f(\lambda^*\,|\,\mathcal{M}_i)\\[2mm]
&=\; \frac{2\,\rho}{\sqrt{1-\rho^2}}\cdot f(\lambda^*\,|\,\mathcal{M}_i) \tag{80}
\end{aligned}
$$

being $\lambda^* = 2\,\sqrt{1-\rho^2}$.

Here are the results for the remaining three models:

$$
\begin{aligned}
f(\rho\,|\,\mathcal{M}_1) \;&=\; \frac{2\,\rho}{\sqrt{1-\rho^2}}\cdot\frac{1}{\pi\,\sqrt{1-(\lambda^*/2)^2}} \tag{81}\\[2mm]
&=\; \frac{2\,\rho}{\sqrt{1-\rho^2}}\cdot\frac{1}{\pi\,\rho}\\[2mm]
&=\; \frac{2}{\pi}\frac{1}{\sqrt{1-\rho^2}} \tag{82}
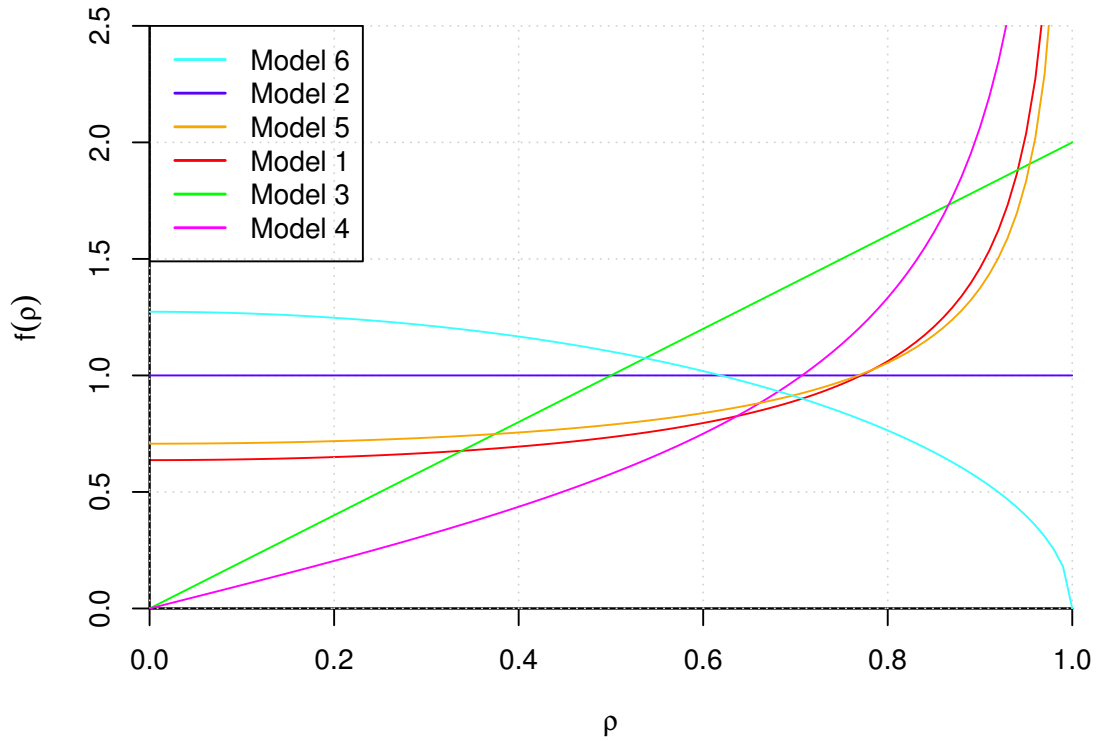\end{aligned}
$$

Figure 24: Distribution of the distance of the chords from center of the circle. (The order of the models in the legend corresponds to the decreasing order of the pdf's for small values of ρ.)

$$f(\rho \,|\, \mathcal{M}_4) \;=\; \frac{2\,\rho}{\sqrt{1-\rho^2}} \cdot \frac{1}{2} \tag{83}$$

$$=\; \frac{\rho}{\sqrt{1-\rho^2}} \tag{84}$$

$$f(\rho \,|\, \mathcal{M}_5) \;=\; \frac{2\,\rho}{\sqrt{1-\rho^2}} \cdot \frac{\sqrt{2 + 2\sqrt{1-(\lambda^*/2)^2}} + \sqrt{2 - 2\sqrt{1-(\lambda^*/2)^2}}}{4 \times 2\sqrt{1-(\lambda^*/2)^2}} \tag{85}$$

$$=\; \frac{2\,\rho}{\sqrt{1-\rho^2}} \cdot \frac{\sqrt{2+2\rho} + \sqrt{2-2\rho}}{8\,\rho} \tag{86}$$

$$=\; \frac{\sqrt{2+2\rho} + \sqrt{2-2\rho}}{4\sqrt{1-\rho^2}} \,, \tag{87}$$

where the form $\sqrt{4 - \lambda^{*2}}$ have been written as $2\sqrt{1-(\lambda^*/2)^2}$, in which we easily recognize 2ρ.

The results are summarized in Fig. 24. As it happened for the pdf's of λ, some of them are divergent for $\rho \to 1$, but it has been checked that all integrals from 0 to 1 are finite and indeed yield 1, as it must be.

Appendix B: A chord length generator based on the Beta distribution

If it makes no sense to talk about the 'correct', or even the 'best', solution in abstract terms, it has also little sense to make an inventory of all possible solutions. The reason on the six ones shown here, let us repeat it again, was the following:

- models $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$ are the 'classical' ones, i.e. those proposed by Bertram himself;

- model $\mathcal{M}_4$ is as simple as the other three (and certainly, in my opinion, much more natural to think about than $\mathcal{M}_3$), justified by the custom of a ruler and compass geometer;

- model $\mathcal{M}_5$ is a variation of $\mathcal{M}_4$, in which the endpoints of the chord are defined by the two intersections of the pencil lead when the compass is rotated clockwise and anticlockwise;

- finally, model $\mathcal{M}_6$ was an attempt to think about of a chord drawing game, in analogy to Buffon's needle.

In all cases (even in the sixth one, that could be indeed played in practice), in order to refer to real cases, the question has been turned into the result of a generator that a 'student' (or anyone else who can write a computer program) might write. And, although at the very beginning one would mainly think about to $\mathcal{M}_1$, and perhaps $\mathcal{M}_2$, once the 'student' starts thinking about the problem and learns how to reuse programming code, we have to be *afraid* (always think to bets!) that more sophisticated extraction models could be implemented.

Finally, once the student realizes that what matters is the distribution of the normalized distance, shown in the previous appendix, the number of methods which can be easily implemented immediately diverges. One only needs an easy generator of a quantity in the range between 0 and 1. The easiest possibility which offers quite some variability in shapes is provided by the Beta distribution, defined as

$$f(x \mid \mathrm{Beta}(r,s)) = \frac{1}{\beta(r,s)} x^{r-1}(1-x)^{s-1} \qquad \begin{cases} r,\, s > 0 \\ 0 \le x \le 1 \, . \end{cases} , \qquad (88)$$

in which the denominator, equal to

$$\beta(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1}\, \mathrm{d}x \, ,$$

defines the *beta function*. Examples of the beta distributions are shown in Fig. 25 (I refrain from showing the resulting chord distributions...). Here it is just how a chord length generator could be implemented in R:

```
> rlchordsBeta <- function(n, r, s) 2*sqrt(1-rbeta(n, r, s)^2)
```

And here it is how to recover very easily the simulations with Methods 2 and 3:

```
> n=100000; hist(rlchordsBeta(n, 1, 1), nc=200, col='blue')
> hist(rlchordsBeta(n, 2, 1), nc=200, col='green')
```

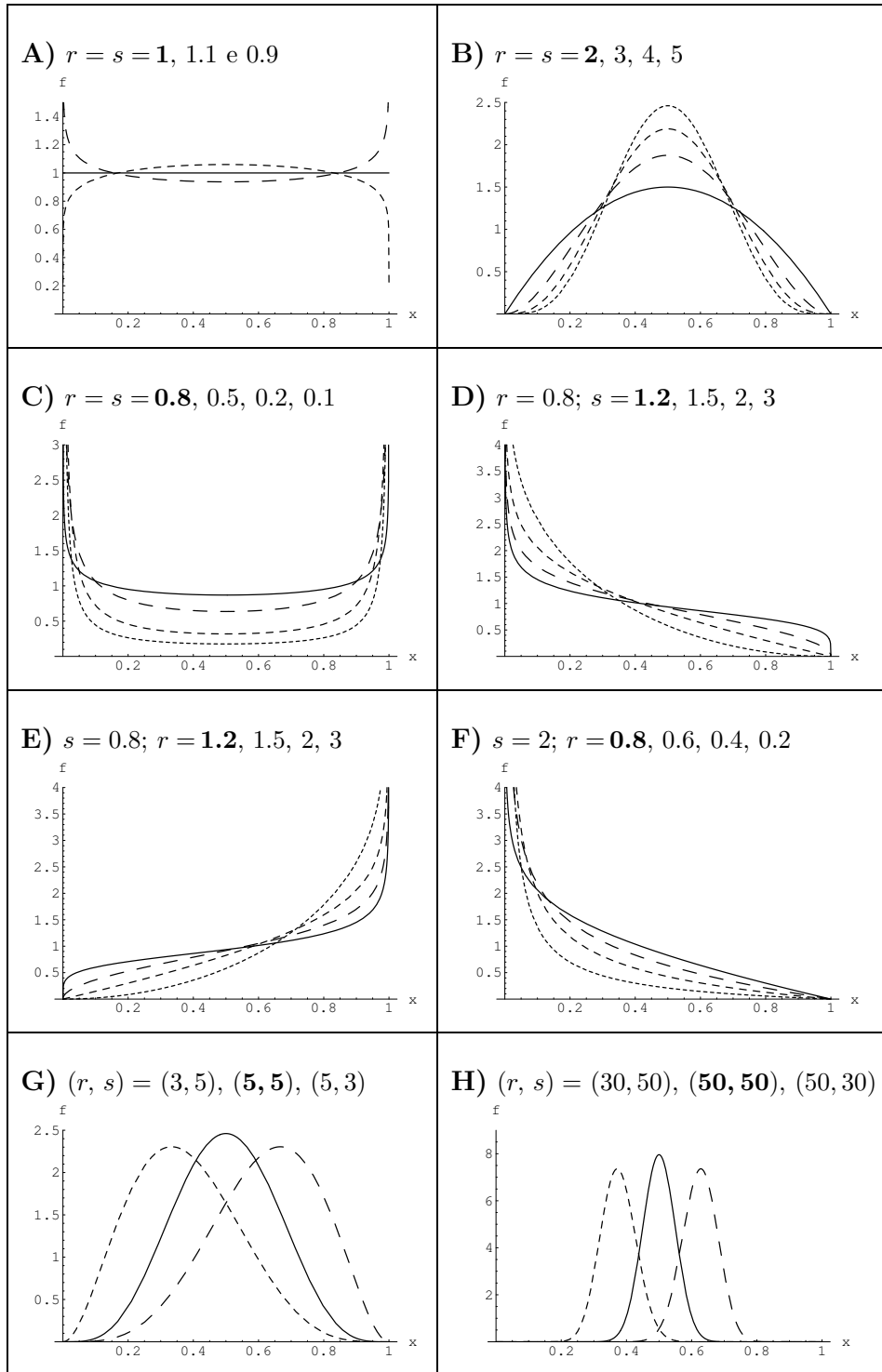Other fancy distributions are left to the fantasy of the reader.

Figure 25: Examples of Beta distributions for some values of r and s. The parameters in bold refer to continuous curves.

44

Appendix C: Turning the Bertrand problem into an inferential-predictive game

> *"I play at écarté with a gentleman whom I know to be perfectly honest.*
> *What is the chance that he turns up the king?*
> *It is 1/8. This is a problem of the* probability of effects.
> *. . .*
> *I play with a gentleman whom I do not know.*
> *He has dealt ten times, and he has turned the king up six times.*
> *What is the chance that he is a sharper?*
> *This is a problem in the* probability of causes.
> *It may be said that it is the essential problem of the experimental method"*
> (Henry Poincaré)

The Bertrand problem can be rephrased in the terms of the causes and the effects of the above quote. If we assume a *cause* (a precise random drawing model) we are uncertain about the *effects* (the chord length). But we might also be uncertain about the model. And usually we do not consider *all* possible models equally likely, where "all model" only means "all models of which we can think about". Usually there will be models in which we believe more and models in which we believe less, and we then rank the possibilities with a degree of belief, or probability.

As a consequence, the uncertainty about the occurrence of each possible effect (the length of a chord, rounded somehow) has to take into account the probability of each value of the length, given the model, and the probability of the model itself. This is done using the following well known rule of probability theory:

$$f(\lambda \,|\, I) \;\; = \;\; \sum_i f(\lambda \,|\, \mathcal{M}_i, I) \cdot P(\mathcal{M}_i \,|\, I)\,, \tag{89}$$

where the background information I has been made here explicit to remind that probability is always conditional probability. A similar formula applies to the probability that the length of the chord is smaller than any given value, that is

$$F(\lambda \,|\, I) \;\; = \;\; \sum_i F(\lambda \,|\, \mathcal{M}_i, I) \cdot P(\mathcal{M}_i \,|\, I)\,. \tag{90}$$

In both cases we have a weighted average, with weights equal to the probabilities of the models. Figure 11 shows with dashed lines the case in which the six models $\mathcal{M}_1$-$\mathcal{M}_6$ are *initially* considered equally likely. In particular, for the original Bertrand question we have, in units of the radius,

$$P(\lambda \leq \sqrt{3} \,|\, I) \;\; = \;\; \sum_i P(\lambda \leq \sqrt{3} \,|\, \mathcal{M}_i, I) \cdot P(\mathcal{M}_i \,|\, I)\,. \tag{91}$$

Let us now think to a different question . We assume we know a chord length, generated by one of the possible models. Our problem does not concern any longer the length of that chord, assuming we are happy with the provided precision, but rather what algorithm was used to generate it. A problem that Poincaré would have

classified as *the essential problem of the experimental method* is then to *guess*, in quantitative terms, which model was used.

Let us imagine, for example, that the extracted chord has a length, in our customary units of the radius, of 0.1, that is $\lambda_1 = 0.1$. Analyzing Fig. 11 we tend to exclude with great certainty models 2, 3 and 6, attributing instead the result most likely to $\mathcal{M}_4$. But we would not believe this model much more than $\mathcal{M}_1$ or $\mathcal{M}_5$. If instead we had (very) good reasons, as we have learned, to exclude $\mathcal{M}_4$ and $\mathcal{M}_5$, we were practically sure that $\lambda_1 = 0.1$ came from $\mathcal{M}_1$. As a consequence, our expectation about the length of a second chord will be that resulting from model $\mathcal{M}_1$ preferred by the *experimental evidence.*

Imagine instead the 'curious' (but not unlikely) situation in which: model 6 was for some reasons excluded; we considered the other five models initially equally likely; the result was $\lambda_2 = 1.4$. In this case the experimental information would practically irrelevant, and this data will not update our opinion concerning the drawing model (see Fig. 11 to figure out the reason).

If we continue the extractions (always using the same generator!) we keep updating the probability of the different models and the probability density function of the length of the *next* chord extracted, or, to make the game simpler, the probability that the next λ will be equal or smaller than $\sqrt{3}$, that we recalculate each time from Eq. (91). Put in this terms the problem is similar to the six box one discussed in Ref. [8], each box having a different content of black and white balls. Indeed we have even the same number of 'causes'. The main difference is that, instead of having only two outcomes (black and white), we have now all possible values between 0 and 2, although discretized. This discretization yields to another important difference. While in the six box problem some of the causes could eventually be 'falsified', i.e. their probability could become exactly zero (a black ball cannot come from a box containing only white balls, and the other way around), in the chord problem falsification is impossible.[16] Given the similarity of the inferential and predictive games I refer then to [8] for a mini introduction to probabilistic inference needed to analyze our problem.

Probability theory teaches us how to update the odds, i.e. the ratio of probabilities, with a rule that it is convenient to rewrite in our case as

$$\frac{P(\mathcal{M}_i \,|\, \lambda_n, I)}{P(\mathcal{M}_j \,|\, \lambda_n, I)} \;=\; \frac{P(\lambda_n \,|\, \mathcal{M}_i, I)}{P(\lambda_n \,|\, \mathcal{M}_j, I)} \times \frac{P(\mathcal{M}_i \,|\, \lambda_{n-1}, I)}{P(\mathcal{M}_j \,|\, \lambda_{n-1}, I)}, \tag{92}$$

where λ_n is the length at the n-th extraction, starting from $n = 0$, that is 'no extraction' and then $P(\mathcal{M}_i \,|\, \lambda_0, I)$ stand for the *priors*. As we see in (92), the *posterior* from the n-th inference becomes the prior of the $(n+1)$-th. Moreover, we are not considering the pdf of λ, but the probability to obtaining a number from our chord generator, rounded somehow. For example, if we round at two significant digits and

---

[16]Some of the pdf's go to zero for $\lambda \to 0$. However the probability in a finite interval around zero is different from zero because the pdf are larger than zero for $\lambda > 0$.

we get $\lambda_1 = 1.23$, the probabilities of interest will be[17]

$$P(\lambda_1 \,|\, \mathcal{M}_i, I) \;=\; \int_{1.225}^{1.235} f(\lambda \,|\, \mathcal{M}_i, I)\, d\lambda \tag{93}$$

$$=\; F(1.235 \,|\, \mathcal{M}_i, I) - F(1.225 \,|\, \mathcal{M}_i, I)\,. \tag{94}$$

In this case we shall make the problem more realistic and avoid the singularities of some of the pdf' (and also be more precise in the calculation of the probability, in the region in which the pdf's are not enough linear).

This is the R code we need to calculate the cumulative distributions for the different models

```
F.lambda <- function(l, mod) {
   if(l<= 0) 0
   else if (l>=2) 1
   else switch(mod,
              2/pi*asin(l/2),
              1 - sqrt(1-l^2/4),
              l^2/4,
              l/2,
              (2+sqrt(2-sqrt(4-l^2))-sqrt(2+sqrt(4-l^2)))/2,
              2/pi*asin(l/2) - l/pi*sqrt(1-l^2/4) )
}
```

(Note how the functions reports 0 for arguments smaller than 0, and 1 for arguments larger than 2).

Then we need a function to calculate the odds between the different hypotheses. We can calculate them with respect to a reference one, from which all others can be easily calculated. Here it is (the default reference model was chosen the one producing a flat distribution of λ, but, as we shall see, the precise default value is irrelevant):

```
chordsModelsBF <- function (l, last.digit=0.01, ref.mod=4) {
  bf <- rep(1,6)
  hld <- last.digit/2
  for (mod in c(1:6)) {
    bf[mod] = (F.lambda(l+hld, mod) - F.lambda(l-hld, mod)) /
              (F.lambda(l+hld, ref.mod) - F.lambda(l-hld, ref.mod))
  }
  return(bf)
}
```

As we see from the list of the arguments, we have to pass also the last digits. The functions return a vector of values. For example
```
> chordsModelsBF(1.23)
```

---

[17]This rule can be applied also at the edges, since the pdf is null outside the range of the variable, as we shall see in the R code, in we shall pay attention that the cumulative function is 0 for λ below 0 and 1 for λ above 2.

[1] 0.8073569 0.7799414 1.2300000 1.0000000 0.8058272 0.6107326

in which we see that this value provides a slight evidence in favor of $\mathcal{M}_3$, as clear from Fig. 11, while 1.95 would favor, in the order, $\mathcal{M}_6$ and $\mathcal{M}_5$, while $\mathcal{M}_4$ is the less favored:

```
> chordsModelsBF(1.95)
```
[1] 2.868580 4.393492 1.950000 1.000000 3.166248 5.454361

Let also try to extreme values, leaving the evaluation of the results to the reader (results are rounded to facilitate the reading).

```
> round(chordsModelsBF(0), 7)
```
[1] 0.6366204 0.0012500 0.0025000 1.0000000 0.5003129 0.0000027
```
> round(chordsModelsBF(2), 1)
```
[1] 18.0 28.3 2.0 1.0 20.0 36.0

At this point we are ready to make the simulations. This is the the rest of the code to make the extractions and to update accordingly the various probabilities:

```
rmod <- sample(1:6)[1]

odds <- matrix(rep(1, 6), c(1,6))
probs <- matrix(rep(1/6, 6), c(1,6))
bf <- matrix(rep(NA, 6), c(1,6))
pLEsqrt3.Mi <- c(2/3, 1/2, 3/4, sqrt(3)/2, (3-sqrt(3))/2, 2/3-sqrt(3)/(2*pi) )
pLEsqrt3 <- sum(probs[1,]*pLEsqrt3.Mi)
fLEsqrt3 <- NA

n <- 200
decimal.digits <- 2
last.digit <- 10^-decimal.digits
l <- NA
for (i in 2:n) {
  l[i] <- round(rlchords(1, rmod), decimal.digits)
  #print(ref.mod)
  bf <- rbind(bf, chordsModelsBF(l[i], last.digit, ref.mod) )
  odds <- rbind(odds, odds[i-1,]*bf[i,])
  ref.mod <- which(odds[i,]==max(odds[i,]))[1]
  odds[i,] = odds[i,] / odds[i,ref.mod]
  probs <- rbind(probs, odds[i, ]/sum(odds[i,]))
  pLEsqrt3[i] <- sum(probs[i,]*pLEsqrt3.Mi)
  fLEsqrt3[i] <- length(l[l<=sqrt(3)])/i
}
```

Then this is how to plot the histories of the probabilities of the models:

```
plot(probs[,1], ylim=c(10^-6,1), xlab='', ylab=expression(P(M[i])),
     log='y', col='red', cex=0.9)
points(probs[,2], col='blue', cex=0.9)
points(probs[,3], col='green', cex=0.9)
```

48

```
points(probs[,4], col='magenta', cex=0.9)
points(probs[,5], col='orange', cex=0.9)
points(probs[,6], col='cyan', cex=0.9)
```

Another interesting plot concerns the probability that the *next* chord will have $\lambda \leq \sqrt{3}$, which was stored, step by step, in the vector `pLEsqrt3`, compared to the relative frequency of occurrence of such a condition in the previous steps:

```
plot(pLEsqrt3, ylim=c(0,1), col='black',
     xlab='n', ylab=expression(P(l<TrS)), cex=0.7)
points(fLEsqrt3, ylim=c(0,1), pch=4, cex=0.7, col='gray')
```

The results are shown in Fig. 26-31, with the models identified by the same color code used in the previous figures. In particular we see that there are model easy to be identified and other more difficult, as it can be understood from the plots in Fig. 11. As far as the comparison between the probability of the next effect calculated from probability theory (black circles) and that evaluated from the past frequency (gray), we see that the former is more stable and rapidly converging to the correct one, corresponding to the model about which we become practically sure.
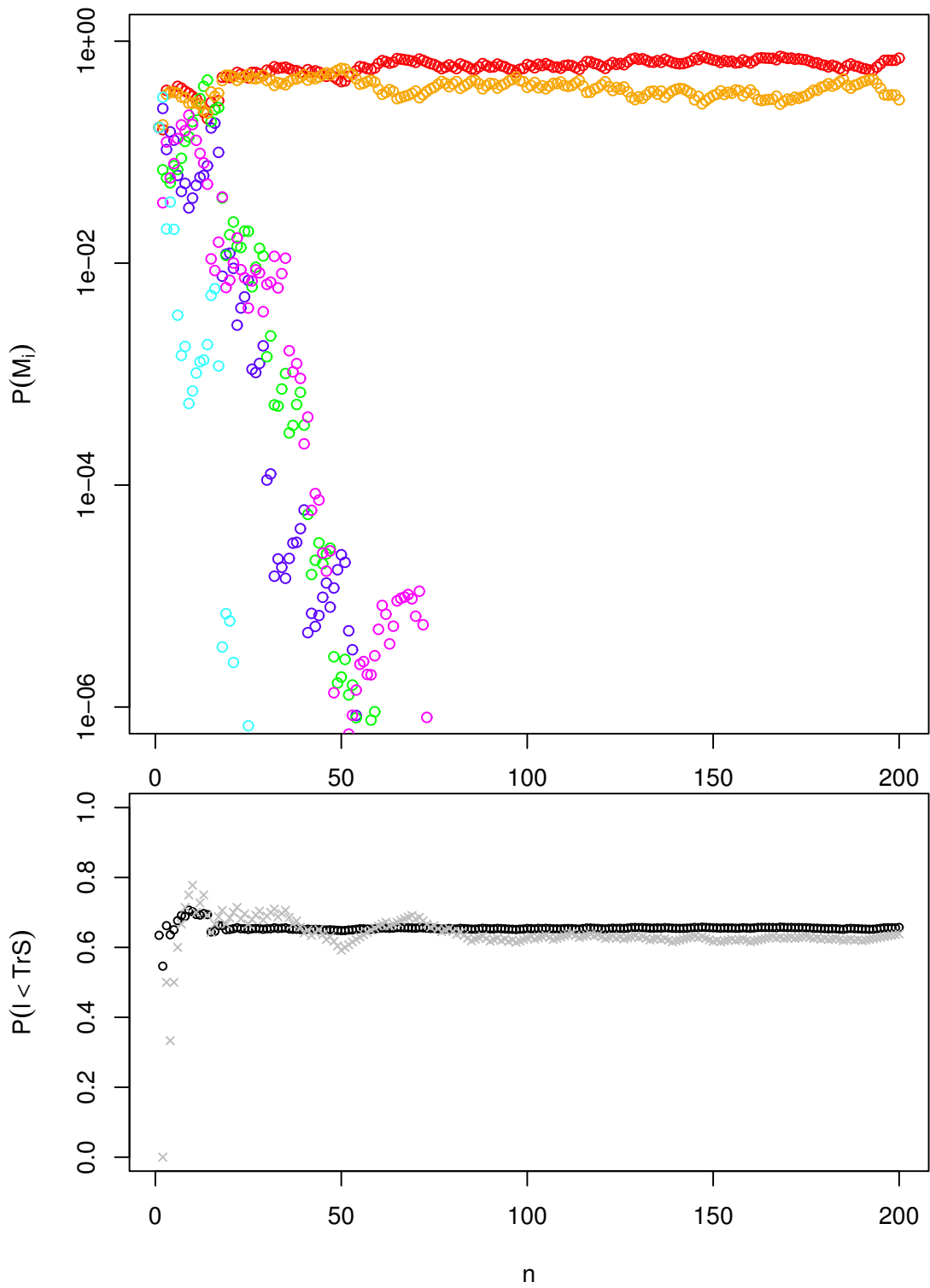
Figure 26: Probability of 'causes' (above) and probability of an 'effect' (below), the last compared with the relative frequency of occurrence (see text). [True model: $\mathcal{M}_1$]
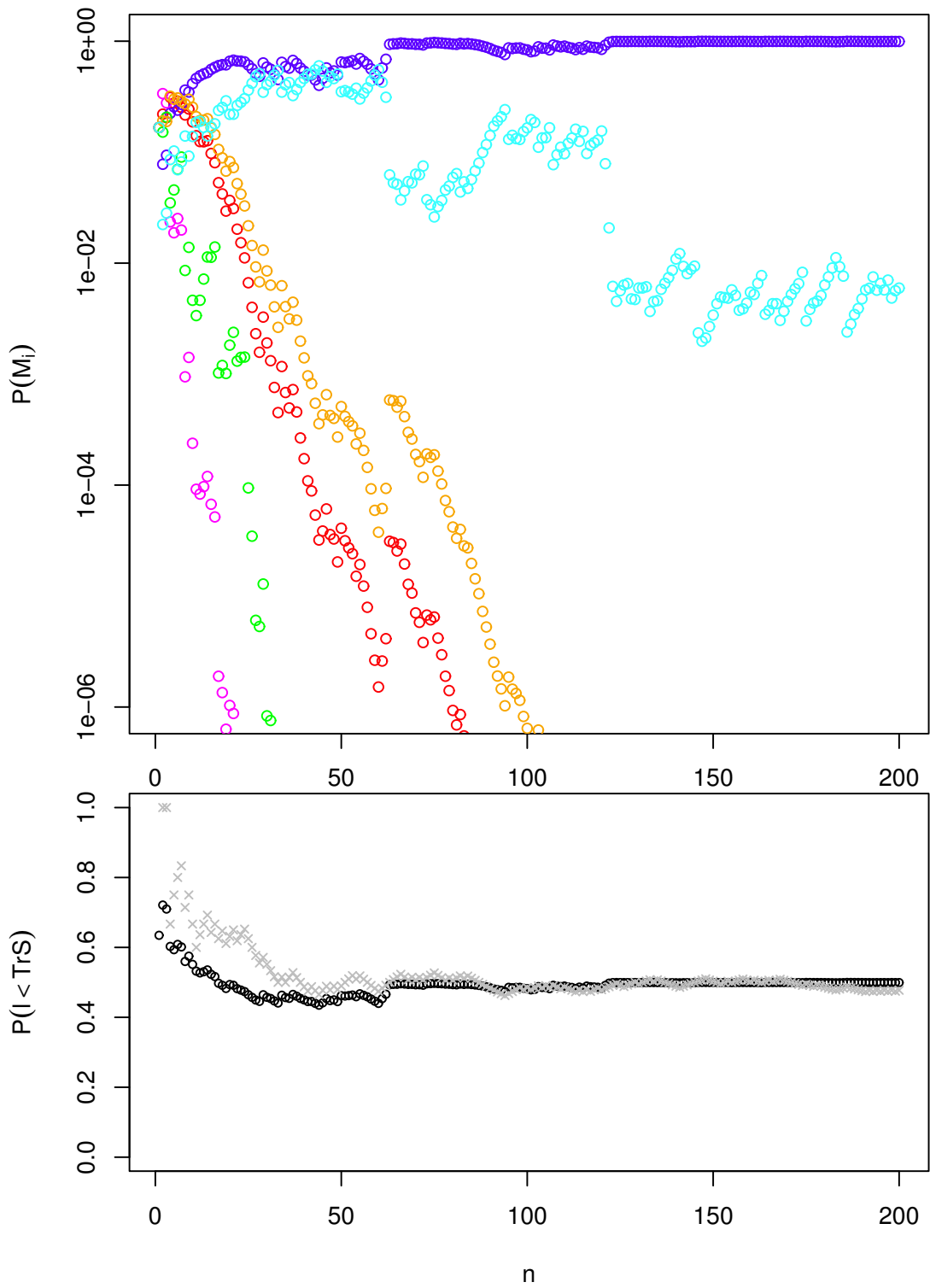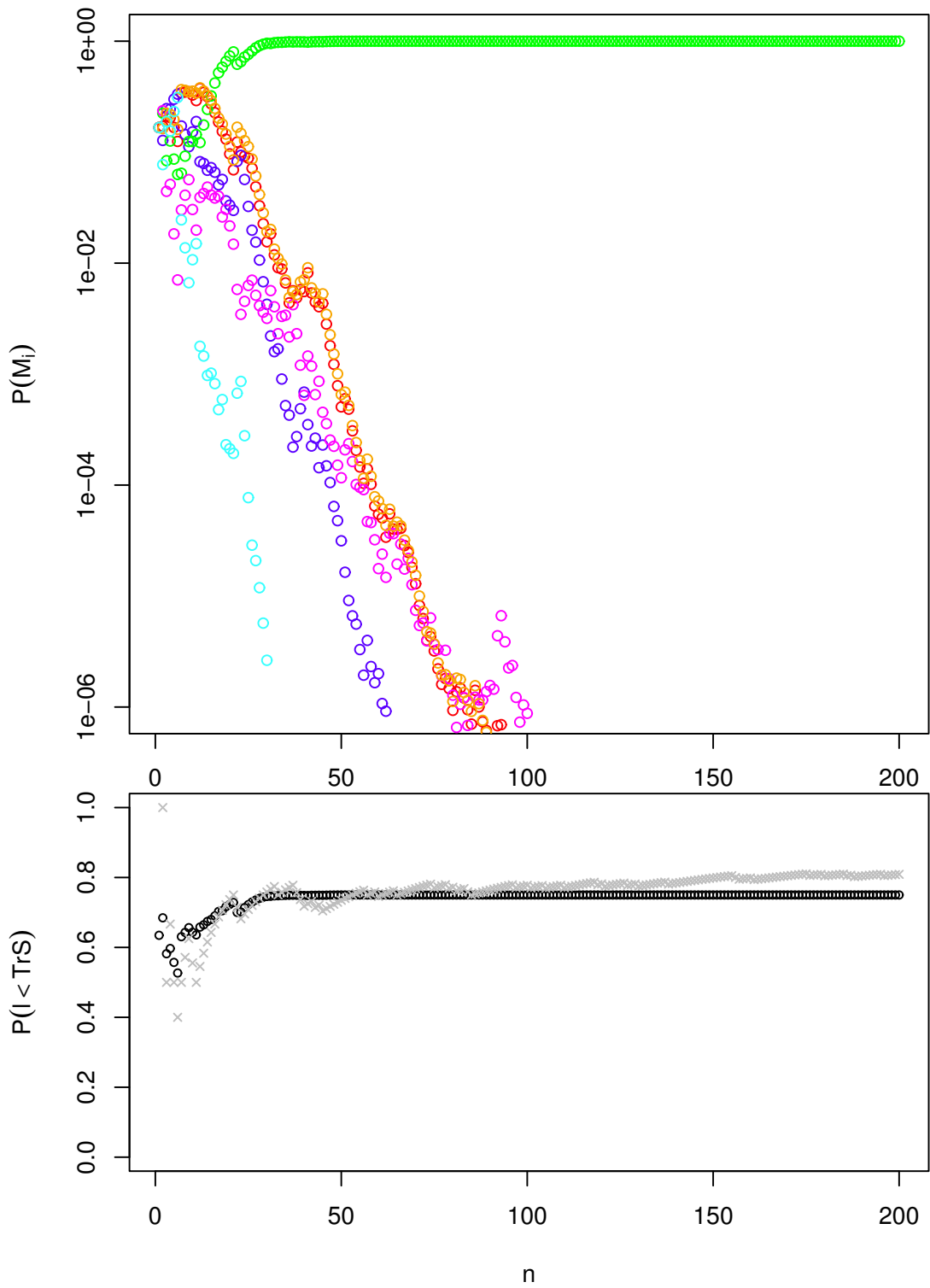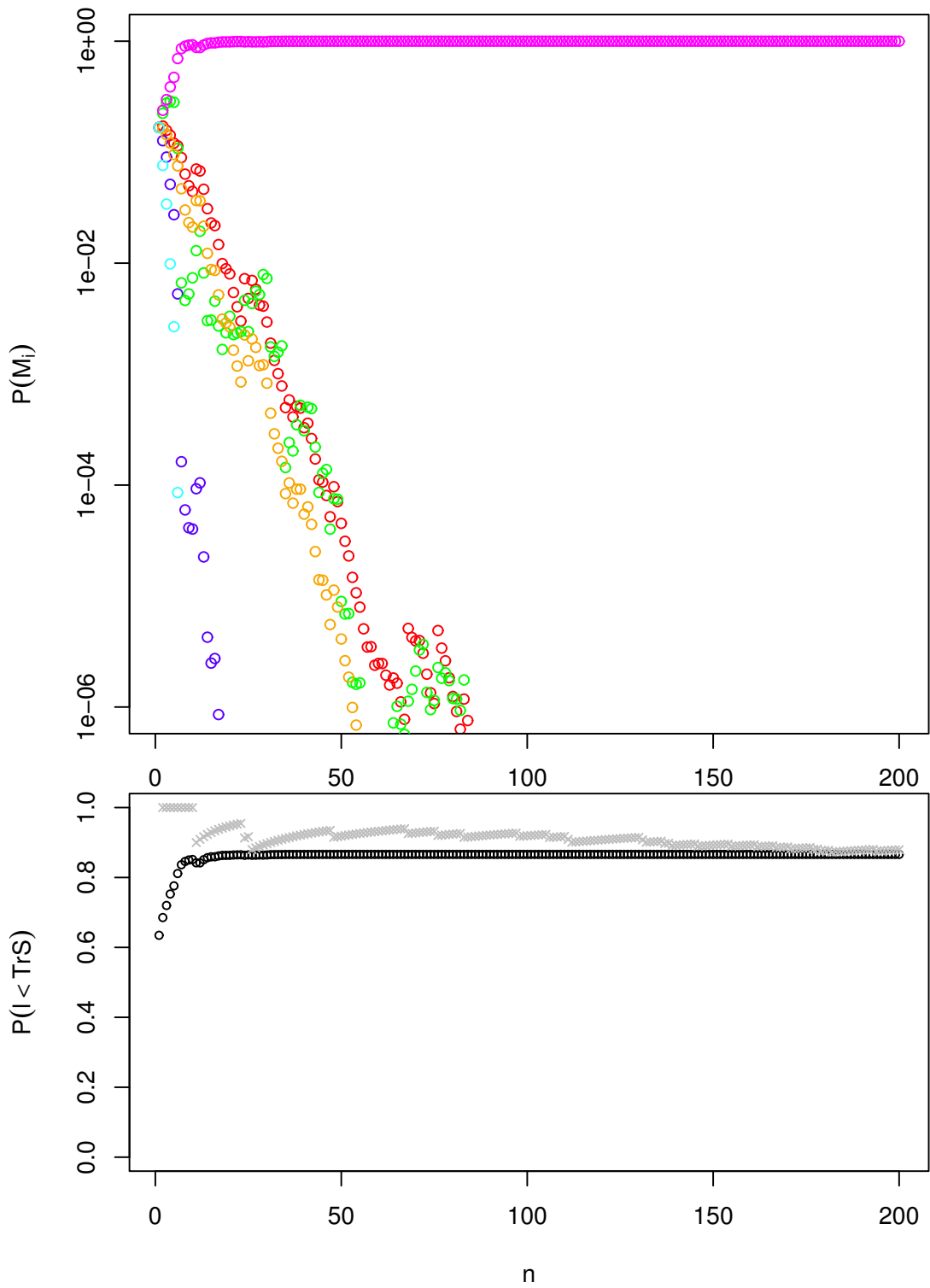
Figure 27: Probability of 'causes' (above) and probability of an 'effect' (below), the last compared with the relative frequency of occurrence (see text). [True model: $\mathcal{M}_2$]

Figure 28: Probability of 'causes' (above) and probability of an 'effect' (below), the last compared with the relative frequency of occurrence (see text). [True model: $\mathcal{M}_3$]

Figure 29: Probability of 'causes' (above) and probability of an 'effect' (below), the last compared with the relative frequency of occurrence (see text). [True model: $\mathcal{M}_4$]
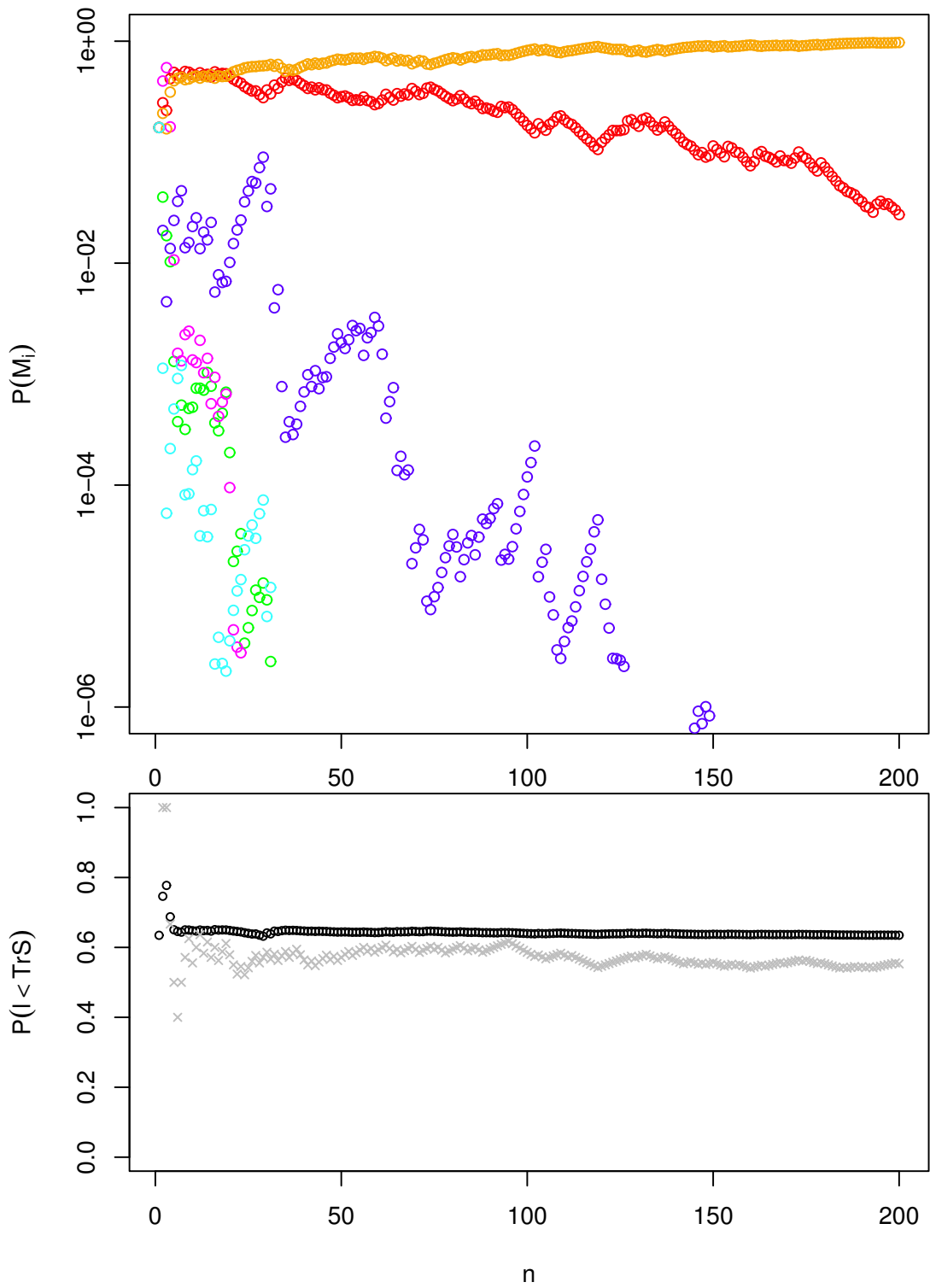
Figure 30: Probability of 'causes' (above) and probability of an 'effect' (below), the last compared with the relative frequency of occurrence (see text). [True model: $\mathcal{M}_5$]
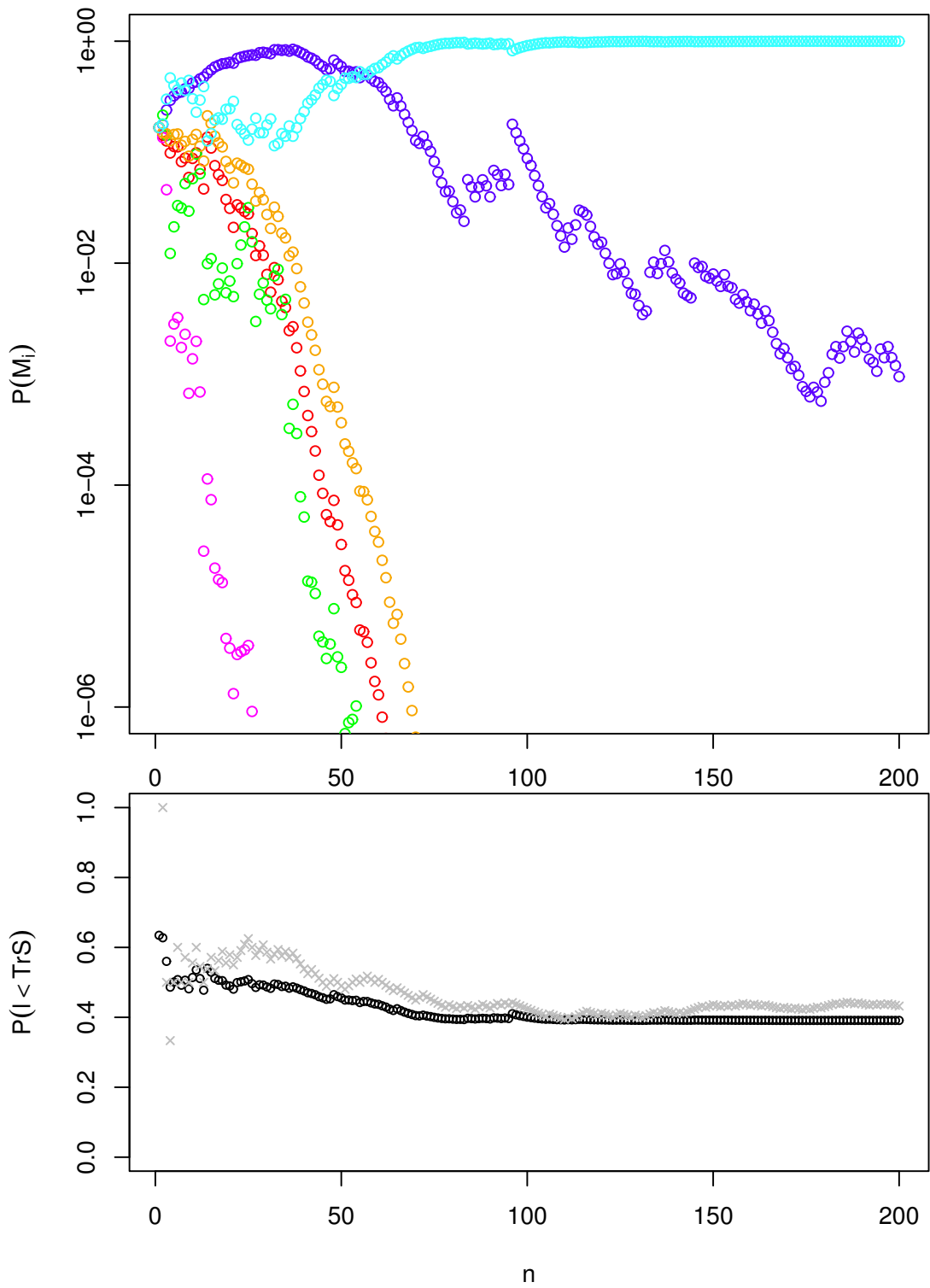
Figure 31: Probability of 'causes' (above) and probability of an 'effect' (below), the last compared with the relative frequency of occurrence (see text). [True model: $\mathcal{M}_6$]