Thou shalt not say “at random” in vain: Bertrand’s paradox exposed

NICOLA CUFARO PETRONI
Department of Mathematics and TIRES, University of Bari
INFN Sezione di Bari
via E. Orabona 4, 70125 Bari, Italy
email: cufaro@ba.infn.it

Abstract

We review the well known Bertrand paradoxes, and we maintain that they do not point to any probabilistic inconsistency, but rather to the risks incurred with a careless use of the locution at random. We claim that these paradoxes spring up also in the discussion of Buffon’s needle, and then that they are related to the definition of (geometrical) probabilities on uncountably infinite sets. A few empirical remarks are finally added.

MSC: 60-01, 60A05, 60E05
PACS: 02.50.Cw, 01.70.+w
Key words: Bertrand’s paradox; Buffon’s needle; Randomness; Change of measure

1 The Bertrand paradoxes

In the first chapter of his classic treatise [1] Joseph Bertrand dwells for a while on the definition of probability, and in particular – in the paragraphs 4-7 – he remarks that the random models with an infinite number of possible results are prone to particularly insidious misunderstandings. He lists then a few examples of problems each admitting equally legitimate, but contradictory answers and suggests then that our questions are ill posed, or more precisely that the required probabilities, based on some at random (au hasard) choice, “sont impossibles à assigner si la question n’est pas précisée davantage” (see [1] p. 7). How it will be made clear later, however – and how it was likely clear to Bertrand himself – the crucial point is less the infinity of the possible outcomes, than their uncountable infinity: a feature shared with

1“L’infini n’est pas un nombre; on ne doit pas, sans explication, l’introduire dans les raisonnements ... Choisir au hasard entre un nombre infini de cas possibles, n’est pas une indication suffisante.” See [1] p. 4
other time honored problems, as for instance that of Buffon’s needle also discussed later in the present paper. We will show indeed that for countably infinite sample spaces the paradoxes do not arise. Bertrand anyhow correctly points out that the proposed questions are fallacious because the locution at random is too careless, but he fails to elaborate further on this point leaving the reader with the odd feeling that something could be inconsistent in a definition of probability. A negligence extended – with few notable exceptions [2, 3] – also to many of the modern textbooks that still bother to mention this topic.

The aim of the present paper is then to address this very point: what are the root and the scope of these seeming inconsistencies? And to accomplish our task we will linger first in the Section 2 on the example that is widely acknowledged today as the paradigmatic Bertrand paradox because its results look especially puzzling. We will then proceed in Section 3 to extend similar remarks to the Buffon needle problem, and in Section 4 to argue that the paradoxes arise only in the event of (geometrical) probabilities defined on uncountably infinite sets. In the last Section 5 we will finally add a few remarks about a possible experimental discrimination among the different legitimate solutions.

2 The circle, the triangle and the chord

Looking at the Figure 1, take at random a chord on the circle \(\Gamma\) of radius 1: what is the probability that its length exceeds that of the edge of an inscribed equilateral triangle (namely exceeds \(\sqrt{3}\))? Three acceptable answers are possible, but they are all numerically different (we always make reference to the Figure 1):

1. To take a chord at random is equivalent to choose the location of its midpoint (its orientation would be an aftermath), and to get the chord longer than the triangle edge it is necessary and sufficient to take this midpoint inside the concentric circle \(\gamma\) with radius \(1/2\) inscribed in the triangle. The required probability is then the ratio between the area \(\pi/4\) of \(\gamma\) and the area \(\pi\) of \(\Gamma\), and consequently we have \(p_1 = 1/4\).

2. By symmetry the position of one chord endpoint along the circle is immaterial to our calculations: then, for a given endpoint, the chord length will only be contingent on the angle (between 0 and \(\pi\)) with the tangent line \(\tau\) in the chosen endpoint. If then we draw the triangle with one vertex in the chosen endpoint, the chord at random will exceed its edge if the angle with the tangent falls between \(\pi/3\) and \(2\pi/3\), and the corresponding probability will be \(p_2 = 1/3\).

---

2See for instance [4] whose final remarks (p. 9) are not really helpful: “We have thus found not one but three different solutions for the same problem! One might remark that these solutions correspond to three different experiments. This is true but not obvious and, in any case, it demonstrates the ambiguities associated with the classical definition, and the need for a clear specification of the outcomes of an experiment and the meaning of the terms ‘possible’ and ‘favorable’.”
3. Always by symmetry, the random chord direction does not affect the required probability. Fix then such a direction, and remark that the chord will exceed $\sqrt{3}$ if its intersection with the orthogonal diameter falls within a distance from the center smaller than $\frac{1}{2}$: this happens with probability $p_3 = \frac{1}{2}$.

We are then left with three different ($\frac{1}{4}$, $\frac{1}{3}$ and $\frac{1}{2}$), but equally acceptable answers. To find the paradox origin we must remember that taking a number at random usually means that this number is uniformly distributed in some domain. It is possible to show however (as also hinted in [2]) that what is considered as uniformly distributed in every single proposed solution can not at the same time be uniformly distributed in the other two: in other words, in our three solutions – by differently choosing what is uniformly distributed – we surreptitiously adopt three different probability distributions, and consequently it is not astonishing at all that the three answers mutually disagree.

To be more precise, let us define (see Figure 1) the three rv pairs representing the coordinates describing the position of our chord in the three proposed solutions:

1. the Cartesian coordinates $(X, Y)$ of the chord middle point

2. the angles $(A, B)$ respectively giving the position of the fixed endpoint and the chord orientation w.r.t. the tangent
3. the polar coordinates \((R, \Theta)\) of the chord-diameter intersection

In every instance however we apparently made the concealed (namely not explicitly acknowledged) hypothesis that the corresponding pair of coordinates is uniformly distributed, but these three assumptions are not mutually consistent, as we will see at once, because they require three different probability measures on the probability space where all our \(rv\)'s are defined. In particular, and by adopting the notation

\[
\chi_{[a,b]}(x) = \begin{cases} 
1, & \text{if } a \leq x \leq b; \\
0, & \text{else}
\end{cases}
\]

the three solutions respectively assume the following uniform, joint distributions (see also Figure 2 for a graphical account of their respective supports):

1. the joint, uniform \(pdf\) on the unit circle in \(\mathbb{R}^2\)

\[
f_{XY}(x,y) = \frac{1}{\pi} \chi_{[0,1]}(x^2 + y^2)
\]

(1)

for the pair \((X, Y)\): here the two \(rv\)'s \(X\) and \(Y\) are not independent

2. the joint, uniform \(pdf\) on the rectangle \([0, 2\pi] \times [0, \pi]\) in \(\mathbb{R}^2\)

\[
f_{AB}(\alpha, \beta) = \frac{1}{2\pi^2} \chi_{[0,2\pi]}(\alpha)\chi_{[0,\pi]}(\beta)
\]

(2)

for the pair \((A, B)\) with independent components

3. and finally the joint, uniform \(pdf\) on the rectangle \([0, 1] \times [-\pi, \pi]\) in \(\mathbb{R}^2\)

\[
f_{R\Theta}(r, \theta) = \frac{1}{2\pi} \chi_{[0,1]}(r)\chi_{[-\pi,\pi]}(\theta)
\]

(3)

for the pair \((R, \Theta)\) again with independent components
Surely enough, if we would adopt a unique probability space for all of our three solutions, the three numerical results would be exactly coincident, but in this case only one of the three \( rv \) pairs could possibly be uniformly distributed, while the other joint distributions should be derived by adopting the well known procedures established for the functions of \( rv \)'s (see for instance [4], Sections 5.2, 6.2 and 6.3). The crucial point here is that there are in fact some precise transformations allowing the change from a pair of \( rv \)'s to the other: by using these transformations we can show now that if a pair is jointly uniform, then the other two can not have the same property.

Without going into the details of every possible combination in our problem, we will confine ourselves to discuss just the relations between the solutions (1) and (3). The transformations between the Cartesian coordinates \(( X, Y )\) and the polar ones \(( R, \Theta )\) are well known:

\[
\begin{align*}
x &= r \cos \theta \\
y &= r \sin \theta \\
r &= \sqrt{x^2 + y^2} \\
\theta &= \arctan \frac{y}{x} \\
\end{align*}
\]

with a Jacobian determinant

\[
J(r, \theta) = \begin{vmatrix}
\frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\
\frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y}
\end{vmatrix} = \begin{vmatrix}
\cos \theta & \sin \theta \\
-\frac{1}{r} \sin \theta & \frac{1}{r} \cos \theta
\end{vmatrix} = \frac{1}{r}
\]

As a consequence (see for instance [4] Section 6.3), if \(( X, Y )\) have the joint uniform pdf (1), then the pair \(( R, \Theta )\) will not be uniform and will have instead the pdf

\[
f_{R,\Theta}^{(1)}(r, \theta) = \frac{r}{\pi} \chi_{[0,1]}(r) \chi_{[-\pi,\pi]}(\theta)
\]

which is apparently different from the \( f_{R,\Theta} \) in (3). By taking advantage of this new distribution \( f_{R,\Theta}^{(1)} \) (coherent now with the choice of a jointly uniform pair \( X, Y \)) it is easy to see that the required probability within the framework of the solution (3) would be

\[
p_3^{(1)} = \int_0^1 \frac{r}{\pi} dr \int_{-\pi}^{\pi} d\theta = \frac{1}{4}
\]

instead of \( p_3 = \frac{1}{2} \), in perfect agreement with the value \( p_1 = \frac{1}{4} \) of the solution (1). Hence the paradox ghosts would daunt us just as long as we unwittingly suppose that in our three solutions the coordinates can all be at once uniformly distributed (hiding that under the careless locution at random), and they will disappear instead as soon as we consistently adopt a unique probability space for all our \( rv \)'s.

### 3 Bertrand vs Buffon

It is interesting to remark now that, while it is known that by the turn of the century several different answers to the Bertrand question were added\(^3\) to the usual three

\(^3\)See for instance [5] quoted in [4]
recalled in the previous section, nobody at our knowledge seem to have noticed that the same kind of paradoxes does in fact appear also in the discussion of the celebrated Buffon needle problem. In its simplest version\footnote{For a more complete discussion see for instance \cite{2} and \cite{7}} a needle of unit length is thrown \textit{at random} on a table where a few parallel lines are drawn at a unit distance: what is the probability that the needle will lie across one of these lines? In the classical answer to this question, since the lines are drawn periodically on the table, it will be enough to study the problem with only two lines by supposing that the needle center does fall between them. The position of the said center along the direction of the parallel lines is also immaterial. The needle position is then defined by just two rv’s: the distance $Z$ of its center from the left line, and the angle $\Theta$ between the needle and a perpendicular to the parallel lines (see (1) in Figure 3). That the needle is thrown \textit{at random} here means that the pair of rv’s $\Theta, Z$ is uniform in $[-\frac{\pi}{2}, \frac{\pi}{2}] \times [0, 1]$, namely that their joint pdf is

\begin{equation}
    f_{\Theta Z}(\theta, z) = \frac{1}{\pi} \chi_{[-\frac{\pi}{2}, \frac{\pi}{2}]}(\theta) \chi_{[0,1]}(z) \tag{4}
\end{equation}

while, with $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$, the needle will lie across a line either when $x \leq \frac{1}{2} \cos \theta$, or when $x \geq 1 - \frac{1}{2} \cos \theta$ (see again (1) in Figure 3). Just by inspecting the pdf (1) in Figure 3 it is then easy to find out that the required probability is

\begin{equation}
    p_1 = \frac{2}{\pi} \int_{-\pi/2}^{\pi/2} \frac{\cos \theta}{2} d\theta = \frac{2}{\pi}
\end{equation}

In the spirit of the Bertrand paradoxes, however, we can give a different answer to the Buffon question (see (2) in Figure 3): the needle position is now identified by looking first to its (vertically) upper end, and by recording its distance $X$ from the
left line. Then we consider where its other (lower) end falls and we mark down its distance \( Y \) from the same left line. If the said left line is in the origin of a horizontal axis, it is apparent that for every value \( 0 \leq x \leq 1 \) of \( X \), the possible values of \( Y \) will be between \( x - 1 \) and \( x + 1 \) (because apparently \( |x - y| \leq 1 \)), and in this framework to throw the needle \textit{at random} will mean that the joint distribution of \( X, Y \) is uniform in the domain shown in (2) of the Figure 4, namely

\[
f_{XY}(x,y) = \frac{1}{2} \chi_{[0,1]}(x) \chi_{[0,1]}(|x - y|)
\]  

(5)

On the other hand it is apparent that for every \( 0 \leq x \leq 1 \) the needle will cross a line when either \( x - 1 \leq y \leq 0 \), or \( 0 \leq y \leq x + 1 \), so that the required probability will correspond to the shaded area in (2) of Figure 4 and hence now \( p_2 = \frac{1}{2} \).

These remarks show then that also the Buffon needle problem is not impervious to paradoxes, and this could be more than just a trifle because of its peculiar experimental status, as argued in the subsequent Section 5.

4 Counting and measuring

It is important to remark that – as already pointed out in the Section 1 – the Bertrand-type paradoxes arise only when we take probability measures on uncountable sets of possible results. To clarify this last point it would be enough to consider another, more simple case among the Bertrand examples: if we ask what is the probability that a real number chosen at random between 0 and 100 is larger than 50, our natural answer is \( \frac{1}{2} \). Since however the real numbers between 0 and 100 are also
bijectively associated to their squares between 0 and 10000, we also instinctively feel that our question should be equivalent to ask for the probability that a random number turns out to be larger than $50^2 = 2500$. If however we take this number at random between 0 and 10000, intuitively again the probability of exceeding 2500 should now be $3/4$ instead of $1/2$. The two problems look equivalent, but their two answers (apparently both legitimate) are different.

Predictably the paradox resolution is similar to that of Section 2: we would readily concede that the probability to exceed 50 for a real number $X$ taken at random in $[0, 100]$ is $p_1 = 1/2$. When however we ask for the probability that $X^2$ taken at random in $[0, 10000]$ exceeds 50$^2 = 2500$, we surreptitiously change our measure by supposing that now $X^2$ is uniform in $[0, 10000]$ and we find $p_2 = 3/4$. But the fact is – as in the previous example – that if $X$ is uniform in $[0, 100]$, then $X^2$ can not be uniform in $[0, 10000]$, and vice-versa. More precisely, if the pdf of $X$ is the uniform

$$f_X(x) = \begin{cases} \frac{1}{100} & \text{for } 0 \leq x \leq 100, \\ 0 & \text{else} \end{cases}$$

then the corresponding, non uniform pdf of $Y = X^2$ is (see again [4] Section 5.2)

$$f_Y(y) = \begin{cases} \frac{1}{200\sqrt{y}} & \text{for } 0 \leq y \leq 10000, \\ 0 & \text{else} \end{cases}$$

and of course the paradox disappears because now, in agreement with $p_1 = 1/2$, it is

$$p_2^{(1)} = \int_0^{2500} f_Y(y) \, dy = \frac{1}{200} \int_0^{2500} \frac{dy}{\sqrt{y}} = \frac{1}{2}$$

It is easy to see however that the paradox does not show up at all when we consider the discrete version of this problem: if we ask for the probability ($p_1 = 1/2$) of choosing at random an integer number larger than 50 among the (equiprobable) integer numbers from 1 to 100, we would in fact recover the same answer ($p_2 = 1/2$) also by asking to calculate the probability of choosing at random a number larger than 2500 among the squared integers $1, 4, 9, \ldots, 10000$, because now our set is again constituted of just 100 equiprobable elements. The crucial difference with the previous continuous version of the problem is that in the case of finitely many (equiprobable) possible results we just enumerate the favorable and the possible items (a situation not changed by squaring the numbers), while for the continuous real numbers we compare the length of the intervals: it is all indeed about the difference between counting and measuring.

This situation, moreover, is not changed even for countably infinite possible results, but for the fact that in this case they can not be equiprobable so that there is not a direct enumeration. For instance the probability that a Poisson distributed integer $N$ exceeds, say, 10 is not changed by asking the probability that $N^2$ exceeds 100, because the countably infinite set of the square integers $1, 4, 9, \ldots$ would be endowed with the same Poisson distribution: the critical feature – common to both
the finite and the countably infinite cases – is now the possibility that every single element be endowed with its own individual non vanishing probability that it also carries with him in every conceivable transformation. A situation totally at variance with that of a geometrical probability on a uncountably infinite set.

5 An almost empiric conclusion

Going back to the example discussed in the Section 2 we have shown that the paradoxes “disappear as soon as we consistently adopt a unique probability space for all our rv’s”, so that – paradoxes notwithstanding – there are no possible formal inconsistencies within our overall probabilistic framework. But this is sheer mathematics, and we are left anyway with three (or more) possible, coherent and perfectly legitimate, probabilistic models giving rise to three numerically different results: which one is true, in the sense that it corresponds to the physical reality? This problem of course can not be solved with a calculation, and should instead be settled – if possible – by comparing the solutions proposed with some empiric result. In other words one should simply perform the experiment of choosing at random a chord in a circle (or something equivalent) in order to compare then its statistics with the calculations: something that to date, at our knowledge, has not been done.

In this vein we add just a few final remarks: first, the possible experiment can not be a simulated one performed on a computer. In this case indeed our experimenter should a priori choose one of the three models to program his computer to produce a particular pair of uniformly distributed coordinates: but in so doing he would have already decided the outcomes of the experiment that coherently will now confirm the chosen model. Second, it is possible that even in some real, physical experiment the outcome can be influenced by the choice of what exactly we decide to measure: different experimental settings could point to different facets of the physical reality, and after all the probability is not listed among the concrete things of this world (see for instance the unconventional viewpoint displayed in [6], vol. 1, Preface) representing rather the state of our information.

We can not refrain however from pointing out in the end that the presence of Bertrand-type paradoxes even in the discussion of Buffon’s needle sheds a different light on this problem: it is known indeed that the classical calculation of the Buffon needle giving the probability $p_1 = \frac{2}{\pi}$ also stimulated several experiments used to get an empirical determination of the approximate value of the number $\pi$ (four of these tests dated from 1850 to 1901 are listed for instance in [2]), and famously known as a first example of the Monte Carlo method. Despite a few reservations about the reliability of these results [2], it is striking that all these four experiments point to a number in the range between 3.14 and 3.16, while our second, proposed alternative solution with $p_2 = \frac{1}{2}$ would require results clustering around 4.00. It is possible – as suggested above – that the quoted results are biased by some unaware bent toward $\pi$, but if confirmed they would suggest that there could be an empirical
meaning in the locution *at random* because, at least in the case of Buffon’s needle, the experiments appear to be able to favor one among several formally legitimate solutions. But it is also apparent that a possible answer would lie outside the reach of this paper, so that for the time being we will stay content with having just clarified the meaning and the scope of the Bertrand paradoxes, by leaving to future inquiries the practical task of empirically deciding among the mathematical models.

**Acknowledgements:** The author would like to thank A. Andrisani, S. Pasazio and C. Sempi for stimulating comments and suggestions.

**References**

[1] *J. Bertrand*, *Calcul des Probabilités* (Gauthier-Villars, Paris 1889)

[2] *B.V. Gnedenko*, *The Theory of Probability* (MIR, Moscow 1978)

[3] *A.M. Mathai*, *An Introduction to Geometrical Probability* (Gordon and Breach, Amsterdam 1999)

[4] *A. Papoulis and S.U. Pillai*, *Probability, Random Variables and Stochastic Processes* (McGraw-Hill, Boston 2002)

[5] *E. Czuber*, *Wahrscheinlichkeitsrechnung und ihre Anwendungen auf Fehlereinschlechterung, Statistik und Lebensversicherung* (Teubner, Leipzig 1903)

[6] *B. De Finetti*, *Theory of Probability* (J. Wiley&Sons, New York 1990)

[7] *A.N. Shiryaev*, *Probability* (Springer, New York 1984)