17/2 (2019), 179–194
DOI: 10.5485/TMCS.2019.0464
tmcs@science.unideb.hu
http://tmcs.math.unideb.hu

Teaching
Mathematics and
Computer Science

Interdisciplinary Secondary-School Workshop: Physics and Statistics

Péter Fejes Tóth, Péter Vankó and Manfred Borovcnik

Abstract. The paper describes a teaching unit of four hours with talented students aged 15-18. The workshop was designed as a problem-based sequence of tasks and was intended to deal with judging dice whether they are regular or loaded. We first introduced the students to the physics of free rotations of rigid bodies to develop the physics background of rolling dice. The highlight of this part was to recognise that cubes made from homogeneous material are the optimal form for six-sided objects leading to equal probabilities of the single faces. Experiments with all five regular bodies would lead to similar results; nevertheless, in our experiments we focused on regular cubes. This reinsures that the participants have their own experience with the context. Then, we studied rolling dice from the probabilistic point of view and–step-by-step–by extending tasks and simulations, we introduced the idea of the chi-squared test interactively with the students. The physics and the statistics part of the paper are largely independent and can be also be read separately. The success of the statistics part is best described by the fact that the students recognised that in some cases of loaded dice, it is easier to detect that property and in other cases one would need many data to make a decision with small error probabilities. A physical examination of the dice under inspection can lead to a quick and correct decision. Yet, such a physical check may fail for some reason. However, a statistical test will always lead to reasonable decision, but may require a large database. Furthermore, especially for smaller datasets, balancing the risk of different types of errors remains a key issue, which is a characteristic feature of statistical testing.

Key words and phrases: Problem-based learning; Chi-squared test, Simulation; Decision rules; Hypothesis test; Free rotation; Rigid bodies .

ZDM Subject Classification: F90, K90, M50, R30.

Introduction

The general debate about statistical inference in statistics education

Statistical inference is known to be difficult. It goes back to the seminal work of R. A. Fisher (Fisher, 1935/1971) and Jerzy Neyman and Egon Pearson (Neyman, 1950, 1952).

Yet, as the curriculum in probability and descriptive statistics would miss the final purpose of the methods developed and thus the students would get a biased perception of probability, some argue that statistical inference should be taught and the research should focus on simplified learning paths through the topic to make sense of the methods in an understandable way.

Borovcnik (1996) proposes simulation, non-parametric methods, and resampling methods for simplifying the teaching of methods and the logic of statistical inference. This idea was also taken up by Cobb (2007), which led to a new strand of research in statistics education, which may be called "Informal Inference". Biehler (2014) discussed the difficulty of informal ways of teaching statistical inference as there are distinct schools of statistical inference and they are related to diverging interpretations of probability. Borovcnik (2017) reminded the community to reconsider some drawbacks of a pure informal inference approach as it is advocated, e.g., in Garfield and Ben-Zvi (2008), or Harradine et al. (2011). Burrill and Biehler (2011) summarise fundamental statistical ideas in the school curriculum but miss all considerations about the so-called Bayesian controversy (Barnard, 1967) on statistical inference. There has been a great and fierce debate on teaching statistical inference around the late 1990's. Moore (1997a, 1997b) made a plea for teaching the significance test (according to R. A. Fisher) while Lindley (1997) and Albert (1997a, 1997b) pleaded for including Bayesian elements. In this line of educational debate, Vancsó (2009; 2013) developed ideas for a parallel approach of both schools of statistical inference in order to make the restriction of each more clearly visible.

Batanero and Borovcnik (2016) describe in details examples of an informal approach towards statistical inference, which resume both Fisher's significance test and the test policy of Neyman and Pearson (and introduce also Bayesian ideas).

The way, statistical tests are introduced here, orientates on the significance test and approaches such as Borovcnik (2014) but at the same time develop the situation towards the Neyman-and-Pearson test policy with considerations of statistical errors of two types (see also Borovcnik, 2015). The approach is similarly decision oriented as suggestions that go back to Riemer (1991) who experimented with a variety of dice that are visibly distinct from the usual cubic form.

The design of the workshop

Inferential statistics is not part of the secondary-school curriculum in Hungary. Present plans to reform the national curriculum include the introduction of basic inferential statistics. Yet, these plans are directed only to those students who aim to achieve an advanced level graduation in Mathematics. Currently, a Working Group of the Hungarian Academy of Sciences (HAS) is exploring the possibilities of such a reform with the additional aim to design feasible learning paths and tasks. One of the main obstacles of introducing inferential statistics to high-school students is that inferential statistics has been part of teachers' education at university level only for around a decade; before that, students met this topic very briefly along their studies. Teachers study probability at the university since 1959, a minor part of which addresses issues on hypothesis testing. Descriptive statistics is included in the pre-service education since the early 2000's though rather from a didactic than from a mathematical point of view. Also, since the topic is not part of the high-school curricula, teachers generally lack experience in actually teaching it.

Because of teachers' background, the aim of the present HAS project is to design a way of introducing inferential statistics in high-school curricula acceptable both for students and their teachers. However, according to preliminary, informal inquiries, teachers can be expected to have a highly dismissive attitude towards the introduction of the new material only because they are unfamiliar with the topic (not counting the beliefs about the mathematical adequacies or difficulties of the field). This also makes it clear that one of the main issues will be in-service training of teachers.

One of the first steps in designing new material was a *workshop* held for students from age 15 to 18, to investigate how difficult it is for them to understand the basics of hypothesis testing. This workshop was an afternoon course for secondary-school students and teachers interested in the topic; participation was on a voluntary basis (Vankó, 2004). The workshop was jointly held by a physics and mathematics teacher. Both parts of the workshop had the same significance. The interaction of the two parts is very important, even though the presented physical description can not explain all the details of the statistical behaviour. With the interdisciplinarity of the workshop we also tried to attract students who were interested only in one of these subjects and make them interested in the other one as well.

During the workshop a key problem on identifying irregular dice was introduced and discussed. We chose this problem as it can be approached from a physical as well a statistical point of view. We used regular and loaded dice where the physical aspect was the investigation of the movement of the die while the statistical aspect was how one can decide whether the die under scrutiny is regular or loaded - a well-known problem in statistics. Neither the physics nor the statistical aspects are part of the curriculum. The dice used in the course were fabricated by a 3D printer. The graphs, statistical calculations, and simulations were performed by Excel. The issues were first discussed from the point of view of physics, then from that of mathematics.

A great advantage of a workshop of this type is that it provides enough time to learn about the theoretical background; it also leaves space to perform several experiments and software simulations. In this paper, we discuss the experiences of the workshop.

Physical background: Rotating rigid bodies

The roll of dice is a sophisticated motion: free rotation in the air, then partly flexible collisions, and landing and rolling on the ground before it stops with one side face up. In case of regular dice, this complex motion guarantees that – independently of how it was rolled – every side should turn face up with the same probability. But in case of loaded dice, the motion is influenced by the physical properties of the dice and the different sides attain different probabilities face up. We might have the *feeling* that the easiest way to load dice is to make one side heavier: then the opposite side will be face up with higher probability. But there are some more sophisticated ways to load dice: for example, one can influence the free rotation of the roll, too. In the workshop, the question "why a cube is used?" was answered by an investigation of free rotation of rigid bodies.

To understand the *free* rotation of a rigid body first let us see a body rotating around a fixed axis. On the left side of *Figure 1*, we can see the simplest model of a statically unbalanced body. The center of mass makes a circular motion and the counterforce pushes both bearings (b_1 and b_2) in one direction. On the right side of *Figure 1*, the center of mass of the rotating body does not move; the resultant force is zero. But the couple has a torque, which pushes the bearings b_1 and b_2 in opposite directions. This is the model of a so-called dynamically unbalanced

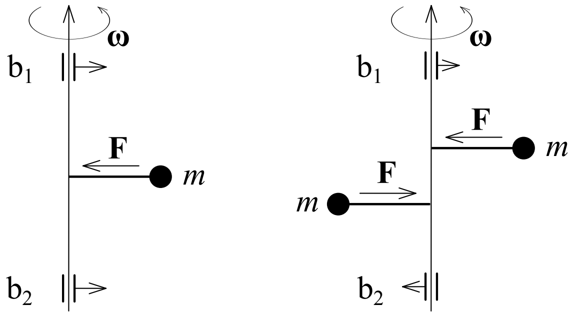

Figure 1. Rotating around a fixed axis

body. (If a wheel of a car was either statically or dynamically unbalanced, the bearings would be ruined, so the wheels have to be balanced with small lead pieces on the rim.) Naturally, such an unbalanced body could not rotate around this axis freely, i.e., without the bearings.

If we see a general rigid body rotating around an axis, the *angular momentum* $\mathbf{L}$ of the body can be determined by summarizing the angular momentum contributions $\Delta\mathbf{L}_i$ of every small Δm_i part of the body, as it is shown on the left side of *Figure 2*.

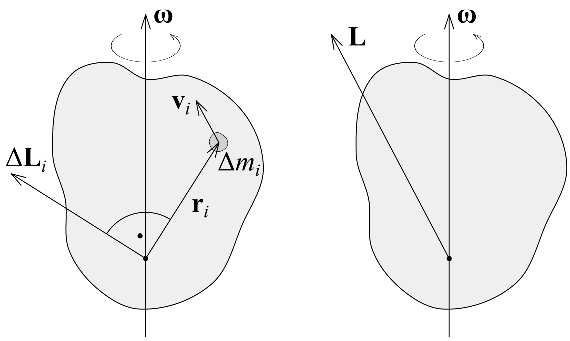

Figure 2. Angular moment of a general rigid body

$$\Delta\mathbf{L}_i = \Delta m_i \mathbf{r}_i \times \mathbf{v}_i = \Delta m_i \mathbf{r}_i \times (\boldsymbol{\omega} \times \mathbf{r}_i)\ , \qquad \mathbf{L} = \sum \Delta\mathbf{L}_i$$

For a general rigid body, the angular momentum vector is usually *not* parallel to the angular velocity vector; this is shown on the right side of *Figure 2*.

The angular momentum vector rotates together with the body around the axis, which implies forces in the bearings. So this body, similarly to the simple dynamically unbalanced model, could not rotate around this axis *freely*, i.e., without the bearings.

The angular momentum vector of the body can be written in the following form:

$$\mathbf{L} = \mathbf{I}\boldsymbol{\omega}\,,$$

where $\mathbf{I}$ is the *inertia tensor*, which is determined by the mass distribution of the body:

$$\mathbf{I} = \begin{vmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{vmatrix} .$$

It can be shown that for every body, there exists a co-ordinate system where the inertia tensor is *diagonal*:

$$\mathbf{I} = \begin{vmatrix} I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_3 \end{vmatrix} .$$

The axes of this co-ordinate system are the principal axes of the body, and I_1, I_2 and I_3 are the principal moments of inertia. If the body rotates around one of these axes, the angular momentum vector is parallel to the angular velocity vector, so it *could be* a free rotation. But from stability reasons, the free rotation is possible only around the axes with the largest (more stable) and the smallest (less stable) principal moments of inertia (and is not possible around the axis with the medium principal moment of inertia).

In the workshop we demonstrated two simple experiments to show that rotating bodies (a rod and a chain) ”chose” the axis with largest moment of inertia (Härtlein, 2011a, 2011b).

The principal axes and the principal moments of inertia determine the *inertial ellipsoid* of the body. The semi-principal axes of the ellipsoid are:

$$a = \frac{1}{\sqrt{I_1}}\,, \quad b = \frac{1}{\sqrt{I_2}}\,, \quad c = \frac{1}{\sqrt{I_3}}\,.$$

The moment of inertia I around a general axis is determined by the ellipsoid:

$$r = \frac{1}{\sqrt{I}}\,,$$

where r is the distance of the surface of the ellipsoid from its center along the direction of the axis.

Let us investigate a die, which has the shape of a cube! It is relatively easy to calculate the moments of inertia around the axes through the face centres. For symmetry reasons they are equivalent: $I_1 = I_2 = I_3 = \frac{1}{6}ma^2$ (where m is the mass of the cube and a is the edge length). It means that the inertial ellipsoid of a cube is a *sphere*, and thus the moments of inertia around *all* axes (through the center of the cube) are equivalent. The moment of inertia of a cube has spherical symmetry, and so the cube can rotate freely around any axis through the center of mass.

Therefore, a *regular* cube is an optimal dice: it rotates randomly around every axis, there are no preferred rotations. In contrast, a cuboid, for example, has a preferred rotation axis (where the moment of inertia is the largest), which influences how the cuboid lands. Similarly, a weighted die has preferred axes, too, and so the different sides will be face up with different probability.

Hypothesis testing: Are these dice regular?

As the participating students have never learnt about hypothesis testing before, we provided a general introduction about what hypothesis testing can be used for. Agriculture, quality control, pharmaceutical, and sociological research, for example, in all these fields hypothesis testing is used regularly. By examples from different professional fields, student realise that this is something that they probably will use later on in their profession.

We chose to use the goodness-of-fit test (chi-squared test) to show how hypothesis testing works. This is a well-known test that is easy to understand, and as such is taught at secondary-school level in countries including inferential statistics in their high-school curriculum. Also, it fits to our task very well: it is easy to experiment with loaded dice even in a classroom and it is easy to simulate such experiments on a computer. We started by explaining the basics of hypothesis testing using the problem of identifying the regularity of dice. We explained how one can make correct or incorrect decisions and demonstrated the difference between type-I and type-II errors (see *Table 1*).

<table>
<tr><td colspan="2"></td><td colspan="2">Our decision</td></tr>
<tr><td colspan="2"></td><td>We say
“Die is regular”</td><td>We say
“Die is loaded”</td></tr>
<tr><td rowspan="2">Real world (truth)</td><td>Null hypothesis:
Die is regular</td><td>Correct decision</td><td>Type-I error (α),
we incorrectly reject the null hypothesis</td></tr>
<tr><td>Alternative hypothesis:
Die is loaded</td><td>Type-II error (β),
we incorrectly stay with the null hypothesis</td><td>Correct decision</td></tr>
</table>

Table 1. Classification of decisions in hypothesis testing

In *Table 1*, α and β are the conditional probabilities of type-I and type-II errors:

$$\alpha = P(\text{type-I error}) = P(\text{we say “The die is loaded”} \mid \text{it is a regular die})$$
$$\beta = P(\text{type-II error}) = P(\text{we say “The die is regular”} \mid \text{it is a loaded die})$$

To start with the goodness-of-fit test, we discuss the implication of the assumption that it is a regular die (null hypothesis); we would expect equal frequencies for each of the faces. Yet, we observe frequencies f_i, which are different from expectation. In the class, two different experiments were discussed whether the die under inspection is regular or loaded; the results are in *Table 2.*

	1	2	3	4	5	6
expected frequency	200	200	200	200	200	200
observed frequency 1	195	210	190	204	205	196
observed frequency 2	170	210	176	202	220	222

Table 2. Frequency table with data discussed

If the expected frequencies are very different from the observed frequencies, we should suspect that the die is regular. The question we have to deal with is “what does very different mean?” This question was raised by students, and that made it very convenient to start to talk about the goodness-of-fit test.

Before we presented the χ^2-criterion, we illustrated the situation by a generic table of observed and expected frequencies (see *Table 3*); note that for the null hypothesis all probabilities are equal, i.e., $p_i = 1/6$.

	1	2	3	4	5	6
expected frequency	np_1	np_2	np_3	np_4	np_5	np_6
observed frequency	f_1	f_2	f_3	f_4	f_5	f_6

Table 3. Generic frequency table

The discrepancy measure of observed frequencies from expected ones under the null hypothesis is:

$$\chi^2_{\text{observed}} = \sum_{i=1}^{r} \frac{(f_i - np_i)^2}{np_i}$$

.

This measure was motivated by a reference to squared distances, which are often used in statistics; only here the denominator is modified and reflects somehow a relative deviation. We can explain to students who are more proficient in mathematics that χ^2_{observed} is a value of a random variable with a sampling distribution approximated by the chi-squared distribution with $r - 1$ degrees of freedom. However, most students do not have the necessary background to understand what a random variable or a distribution function is. Such gaps in the mathematical representation are very frequent in statistics as the mathematical relations are complex.

In such cases, in teaching we have to fill the gaps by analogy or refer to the *phrase*: "mathematicians can prove that". This type of "argument" is often used in countries where the goodness-of-fit test is part of the secondary-school curriculum (Blythe et al., 2012, p. 233) and even in higher-education textbooks it is unusual to go deeper into the mathematical background (Walpole et al., 2012, p. 372) – usually, it is only taught how you can make a decision after obtaining the chi-squared value.

As we had motivated and well-equipped students in the workshop, we discussed a little about distributions in general and we explained that if the die is regular (null hypothesis, H_0 is true), then χ^2_{observed} is a value of a random variable, which can be approximated very closely by the chi-squared distribution with a 5 degrees of freedom (*Figure 3*).

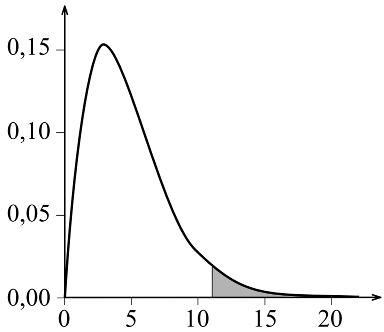

Figure 3. χ^2 distribution with 5 degrees of freedom

α	χ^2_{critical}
0.100	9.2
0.050	11.1
0.010	15.1
0.001	20.5

Table 4. Critical values for χ^2-test with 5 degrees of freedom

At the end of this session, we explained (and showed on the graph of the chi-squared distribution) how unlikely it is that χ^2_{observed} is greater than a certain critical number (which depends on the probability of Type I error α), which is based on the null hypothesis H_0 being true (in *Table 4* various α are shown with the critical value related to it for 5 degrees of freedom).

Thus, making a decision is quite easy:

- We accept the null hypothesis H_0 if $\chi^2_{\text{observed}} \leq \chi^2_{\text{critical}}$; in this case we say that the data complies with the assumption of the regularity of the die.
- We reject the null hypothesis H_0 and accept the alternative hypothesis H_1 if $\chi^2_{\text{observed}} > \chi^2_{\text{critical}}$. this does not mean that we have proven that the die *is* irregular; in this case, we just know that if the die is regular, the probability of getting such a high value for χ^2 is very small (it is less then α).

It is necessary to point out that in both cases there is a risk (chance) of making a wrong decision. In the first case, when we accept the null hypothesis (H_0) but in fact the alternative hypothesis (H_1) applies, we commit a type-II error; however as we do not know, which specific distribution applies for the die under scrutiny, we can not calculate what the probability of that is. In the second case, when we reject the null hypothesis H_0, we do know what the probability of making a type-I error is, as we have preselected α for our decision rule. This is how hypothesis testing works.

After providing this theoretical background, we actually played with dice that were either loaded or not. We formed eight groups (pairs of students). The aim of the game was to ground the basic concepts and to experience the notions, the logic, and the method of hypothesis testing by practical activities. Though this method was introduced earlier (Lawton, 2009; Dambolena et al., 2009), our approach places greater emphasis on the hands-on in-class experiment.

Each of the groups should find out whether their die was regular. Initially, there was a fierce discussion about how many times the die should be thrown in order to obtain sufficient evidence for the decision. After a while, there was a compromise of 200 throws as it seemed convenient to have the same number of throws in each group to facilitate the comparison of the results between the groups. Each group recorded the results in a frequency table and later the data sets were represented by bar charts.

Based solely on their data, they should decide whether their die is regular or not. After each group had made their decision, they exchanged data and decisions with the other groups. A class discussion followed about the quality of their decision. The decision led to two groups changing their prior decision. Finally, they were asked to perform the chi-squared test on their data. One further decision (rejection) was not "confirmed" by the chi-squared result, so it had to be changed. Remarkably, only one group had a significant chi-squared test rejecting the null hypothesis of the regularity of the die. When the students reflected whether they should change their decision in the light of the information of the other groups, the discussion about the types of errors that can be committed was revisited.

In the next unit, we used Excel with its random number generator to simulate the scenario under the null hypothesis of a regular die. The software is more flexible as it allows generating larger datasets. The impact of increasing sample size was investigated. The students should experience that the pattern of the observed frequencies (relative frequencies) stabilises with larger sample size but can be erratic with smaller sample size. This general pattern can be illustrated best by repeating the whole scenario of simulation looking on the corresponding bar charts like one would look on an animated film.

Some examples, which we generated in the workshop with a random number generator assuming a regular die, are shown in *Figure 4*.

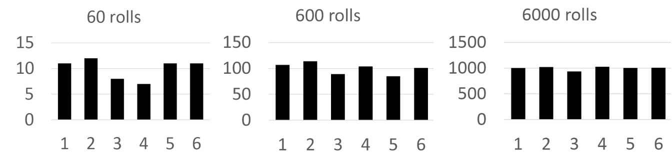

Figure 4. Histogram of a regular die rolled 60, 600 and 6000 times

With 60 rolls of a regular die, the histogram does not resemble a uniform distribution (*Figure 4*, left); even in case of 600 rolls it looked uneven (middle). However, rolling 6000 times seemed to reproduce the expected results (right). Repetition of the scenario showed that the pattern is stable.

We also used simulation of various differently loaded dice. It was interesting to examine, which histogram helped to make good decisions (*Figure 5*).

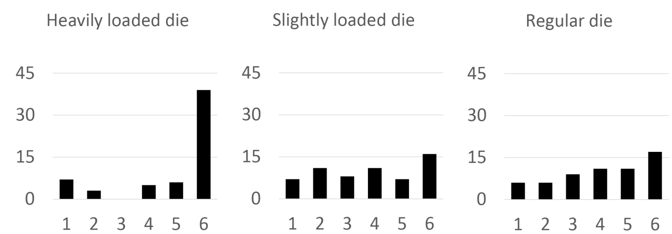

Figure 5. Histogram of differently loaded dice rolled 60 times

A heavily loaded die may easily be identified already after 60 rolls (*Figure 5*, left). On the other hand, neither a lightly loaded, nor a regular die can be recognised so easily (middle and right). Based merely on this data, we easily could come to a wrong decision about regular or slightly loaded die: for the regular die, we may wrongly conclude that we are dealing with a loaded die (type-I error) while for a slightly loaded die, we may wrongly conclude that the die is regular (type-II error). It seemed very helpful, to run a lot of simulations and make the students guess whether the simulated die is regular or not. In this way the students could understand easily when they made a type-I or a type-II error.

Conclusion

In the early phases of the workshop, the students decided about the sample size of 200 in order to judge whether the inspected die is regular or not. This decision has been made with quite a fierce dispute between the students. From the simulation scenarios later in the workshop, they learned that even samples of 600 throws of a regular or a loaded die could easily result in observed frequencies, that lead to a wrong decision (based on the visual expression of the bar chart or the formal chi-squared test). This experience helped the students to recognise that in our case, with moderately weighted dice, the initial decision for 200 as sample size was unsatisfactory: 200 is simply not enough for a proper decision for a small type-I error (less than 0.05) to guarantee a reasonable type-II error to recognise that the die is loaded. As we have demonstrated in our simulations, a heavily loaded die reveals its properties already in a much shorter sequence of rolls and can thus easily be detected by inspecting the relative frequencies.

The various scenarios of loaded dice, which we investigated by simulation, let the students recognise that some cases of a loaded die are easy to detect (which leads to the rejection of the null hypothesis). However, the question how to find a suitable measure how far the distribution of a loaded die is from the uniform distribution of a regular die was not easy for them to answer. Even if they could refer to a chi-squared criterion (which was used earlier for a different purpose of course), they judged the difference of two distributions by the maximum difference of one face. Still we regard it as a positive result of the workshop that the students explicitly referred to the fact that in some cases it is easier to decide whether the die is loaded or not, and in other cases it is more complicated.

One great advantage of the workshop is that the students recognised that in these cases it is much better to have more data to base the decision on. In this way, the students became also aware of the fact that it is possible to decrease the second error for a fixed level of the first error by increasing the database (i.e., by increasing the sample size).

The class discussion touched also the issue that the types of error are antagonistic. That means that making the type-I error smaller causes the type-II error to get larger. Or, being prepared to have a larger type-I error for the decision rule makes the type-II error smaller.

Interestingly, only by the end of the classroom discussion, the students went back to the physics part of the workshop. In the concluding unit, students started to discuss whether there are physical experiments to check if whether the die is loaded or not. That led to the discussion of a ”floating” experiment, which gives yields much faster answers for a specific loaded die than with data analysis could provide but this way of inspection might fail for other types of loaded dice.

In this way, the study of the physics of rigid bodies clarifies that the homogeneous cubic form is the only one that guarantees a uniform distribution for six sides (because of the properties of dynamic rotations). Yet, physics does not provide a solution for detecting loaded dice at least as a simple method for all cases. On the contrary, the method of data analysis neglecting physical interrelations provides a clear decision with guaranteed probability for errors of both types if only enough data is collected.

References

Albert, J. (1997a). Teaching Bayes' rule: A data-oriented approach. *The American Statistician*, *51*(3), 247–253.

Albert, J. (1997b). Reply. *The American Statistician*, *51*(3), 271.

Barnard, G. A. (1967). The Bayesian Controversy in statistical inference. *Journal of the Institute of Actuaries*, *93*(2), 229–269.

Batanero, C., & Borovcnik, M. (2016). *Statistics and probability in high school.* Rotterdam: Sense Publishers.

Biehler, R. (2014). On the delicate relation between informal statistical inference and formal statistical inference. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9).* The Hague: International Statistical Institute and International Association for Statistics Education. Retrieved 04/18/2019, from `http://iase-web.org/Conference_Proceedings.php`

Blythe, P., Fensom, J., Forrest, J., & Waldman de Tokman, P. (2012). *Mathematical Studies Standard Level.* New York: Oxford University Press.

Borovcnik, M. (1996). Trends und Perspektiven in der Stochastik-Didaktik (Trends and perspectives in the didactics of stochastics). In G. Kadunz, H. Kautschitsch, G. Ossimitz, & E. Schneider (Eds.), *Trends und Perspektiven. Schriftenreihe Didaktik der Mathematik* (Vol. 23, pp. 39–60). Vienna: Hölder-Pichler-Tempsky.

Borovcnik, M. (2014). Modelling and experiments–An interactive approach towards stochastics. In T. Wassong, D. Frischemeier, P. Fischer, R. Hochmuth, & P. Bender (Eds.), *Mit Werkzeugen Mathematik und Stochastik lernen.*

Using tools for learning mathematics and statistics (pp. 267–281). Berlin: Springer.

Borovcnik, M. (2015). Risk and decision making: The "logic" of probability. *The Mathematics Enthusiast*, *12*(1,2&3), 113–139.

Borovcnik, M. (2017). Informal inference–Some thoughts to reconsider. In *Proceedings of 61st ISI World Statistics Congress.* The Hague, The Netherlands: International Statistical Institute.

Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education. A Joint ICMI/IASE Study* (pp. 57–69). New York: Springer.

Cobb, G. W. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education*, *1*(1). Retrieved 04/18/2019, from `http://www.escholarship.org/uc/item/6hb3k0nz`

Dambolena, I. G., Eriksen, S. E., & Kopcso, D. P. (2009). An Intuitive Introduction to Hypothesis Testing. *INFORMS Transactions on Education*, *9*(2), 53–62.

Fisher, R. A. (1935/1971). *The design of experiments.* Edinburgh: Oliver & Boyd.

Garfield, J. B., & Ben-Zvi, D. (2008). Learning to reason about statistical inference. In J. B. Garfield & D. Ben-Zvi (Eds.), *Developing students' statistical reasoning* (pp. 261–288). New York: Springer.

Harradine, A., Batanero, C., & Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education. A Joint ICMI/IASE Study* (pp. 235–246). New York: Springer.

Härtlein, K. (2011a). *Unfixed axis I. Video.* Retrieved 04/18/2019, from `https://fizipedia.bme.hu/index.php/Szabad_tengely_I.`

Härtlein, K. (2011b). *Unfixed axis II. Video.* Retrieved 04/18/2019, from `https://fizipedia.bme.hu/index.php/Szabad_tengely_II.`

Lawton, L. (2009). An Exercise for Illustrating the Logic of Hypothesis Testing. *Journal of Statistics Education*, *17*(2), 1–9.

Lindley, D. V. (1997). Discussion. *The American Statistician*, *51*(3), 265–266.

Moore, D. S. (1997a). Bayes for beginners? Some reasons to hesitate. *The American Statistician*, *51*(3), 254–261.

Moore, D. S. (1997b). Reply. *The American Statistician*, *51*(3), 272–274.

Neyman, J. (1950). *First course in probability and statistics.* New York: Henry Holt.

Neyman, J. (1952). *Lectures and conferences on mathematical statistics and probability.* Graduate School, U.S. Dept. of Agriculture: Washington.

Riemer, W. (1991). *Stochastische Probleme aus elementarer Sicht (Stochastic problems from an elementary viewpoint).* Mannheim: Spektrum Akademischer Verlag.

VancsĂł, Ö. (2009). Parallel discussion of classical and Bayesian ways as an introduction to statistical inference. *International Electronic Journal in Mathematics Education*, *4*(3), 291–322.

Vancsó, Ö. (2013). Teaching statistical inference from multiple perspectives integrating diverging schools of inference. In *Proceedings of 59th ISI World Statistics Congress* (pp. 1108–1113). The Hague, The Netherlands: ISI.

Vankó, P. (2004). An Experimental Problem of a Competition Discussed in a Secondary School Workshop. *Physics Competitions*, *6*(1), 45–47.

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2012). *Probability and Statistics for Engineers and Scientists* (9th ed.). Boston: Prentice Hall.

Péter Fejes Tóth
Szent István University
Department of Biometrics and Agricultural Informatics
H-1118 Budapest, Villányi srt. 29–43.

E-mail: fejestothpeter@yahoo.com

Péter Vankó
Budapest University of Technology and Economics
Department of Physics
H-1111 Budapest, Budafoki srt. 8.

E-mail: vanko@eik.bme.hu

Manfred Borovcnik
University of Klagenfurt
Department for Statistics
A-9020 Klagenfurt, Universitätsstrasse 65

E-mail: manfred.borovcnik@aau.at

(Received May, 2019)