Ten Challenges and Explainable Analogs of growth functions and distributions for statistical literacy and fluency

Georges Hattab*
Center for Artificial Intelligence in Public Health Research (ZKI-PH)
Robert Koch-Institute, Nordufer 20, 13353 Berlin, Germany

Abstract

Personalized learning, science education, and public understanding of science are linked in many ways. Studies of public understanding of science suggest that citizen science literacy is not just about reforming school science curricula. Understanding statistics is a challenge for many people, and customized learning methods are required. Statistics are useful for keeping records, calculating probabilities, and providing knowledge. Basically, they help us understand the world a little better through numbers and other quantitative information. Motivated by the overall goal of promoting understanding of statistical concepts and minimizing misinformation, thirteen gamification and data physicalization initiatives help us to offer ten challenges. They allow players to record, update, and explore the characteristics and differences between different statistical concepts. This work combines dynamic physical visualization and gamification to explain linear and exponential functions and a set of distributions using analog explicable cubes as a means of representation.

Index Terms: Mathematics of computing—Probability and statistics; Social and professional topics—User Characteristics—Age; Human-Centered Computing—Visualization;

1 Introduction

Understanding statistics is a challenge for many people. It requires familiarity with statistical concepts of linear and exponential functions and distributions like uniform, Gaussian, Poisson, Exponential, Gamma, Binominal, and Chi-squared. This work explains these functions and distributions by using analog explainable cubes as means of representation. The overarching goal is to boost the understanding of statistical concepts and minimize misinformation by means of gamification [1–3] and data physicalization [4, 5].

As scientists, there is a responsibility and pressing need to better communicate simple concepts and ideas for improved public understanding, numerical and statistical literacy. Analog or physical (real-world) analogies of changes in a data set with million data points have become limited by screen space, resolution, and our estimation of underlying functions and distributions. Using a set of twelve cubes in a bag and a pencil, growth functions and certain distributions are explained in an analog fashion. Moreover, in an effort to address one of the 14 Grand Challenges for Engineering in the twenty-first century from the National Academy of Engineering, this work develops concepts aimed at advancing personalized learning of statistics through physical visualization. Throughout the education system, teaching has traditionally followed a one-size-fits-all approach to learning, with a single set of instructions given in the same way to all students in a given class, regardless of differences in ability or interest. Since there is a wide variety of learners, a variety of methods should be developed to better accommodate different learners. Learners may be highly motivated and self-directed (they learn best by exploring an area of knowledge on their own with little or no guidance), prefer guidance and a more structured approach (they are generally motivated when the topic is interesting), be motivated by external rewards and learn best with step-by-step instruction, or be resistant and unmotivated or interested in achieving goals set by others. Among the basic types of learners, some prefer to learn by example, some by finding answers to questions, and some by solving problems on their own. Under different conditions, people may even change their preferences, preferring examples in some contexts and questions in others. Current efforts in personalized learning have led to personal learning approaches ranging from modules that students can master at their own pace to computer programs designed to tailor the way they present content to the learner's personality. Not surprisingly, many efforts to address this phenomenon involve computer-based instruction, often in the classroom or via the Internet. Among the many projects attempting to address the challenge of personalized education are web-based "intelligent" educational systems [6], the development of "recommender" systems that guide individual learning using web-based resources [7], and the creation of algorithms that tailor recommendations to the student's abilities [8]. This work takes a rather analog approach to provide learners with the means to discover, play, and understand. This approach explores gamification considerations using physical visualizations and can be made available to the general public or policy makers.

*e-mail: georgeshattab@gmail.com

(a) (b)

Figure 1: Sugar cubes (a), wooden cubes (b), or even LEGO® pieces (not shown) may be used

First, thirteen initiatives are highlighted from gamification and data physicalization. Second, we build upon these initiatives to propose ten challenges. They comprise two functions and eight distributions. While linear and exponential growth functions are presented to combat exponential growth bias [9, 10], the collection of distributions includes: uniform, normal or Gaussian, Poisson, gamma, Chi-squared, Bernoulli, binomial, and exponential. Altogether, they make ten challenges.

2 Methods

Thirteen initiatives are being considered to make statistics accessible to the general public. By considering simple tokens, players can grasp the constructive aspect behind each visualization [11]. To create explainable analogs of growth functions and distributions for statistical literacy and fluency, the three motivations of data physicalization are adopted: discover, present, and enjoy. Moreover, ten gamification techniques are used to spark motivation, engage and encourage learning [12, 13]: flow, task, challenge, customize, feature, curiosity, surprise, achievement, competition, and collaboration. They are presented below. Then, each challenge is described sequentially, from easy to hard difficulty.

2.1 Discover

Data physicalization is interested in clear encodings. It combines them with storytelling elements that guide the audience to discover the ideas the designer wants to communicate. Whether it was initially meant to analyze data or to be better at communicating or teaching data insights, this work is inspired by nineteenth century work to explain scientific concepts to students or the public. To discover things, guidance is offered with a sequence of events. Discovery is conceived as an implicit aspect that is linked to the presentation where scientific concepts crystallize.

2.2 Present

Presentation conveys information from a speaker or a designer to an audience. The presentation here means engaging and connecting with an audience with high level information and relatable examples to better understand the statistical concepts.

2.3 Enjoy

Learning new things comes with a certain amount of satisfaction. The purpose of this step is to first provide positive feedback and encouragement, and then to elaborate on previously discovered concepts and present them in a domain-specific manner. In this case, relying on statistical terminology and mathematical notations. The introduction of the latter is done progressively to reinforce the player's knowledge.

2.4 Flow

Achieving what Hungarian-American psychologist Mihaly Csikszentmihalyi called "flow" is possible by using a balanced task as described in 5. Task. This is also possible by relying on a practical, progressive task that offers guidance.

2.5 Task

If a task is too easy, you will be bored. If it's too hard, you'll get frustrated. When creating a game or game experience, you need to find a balance. A task should be challenging, but doable. Unfinished tasks often bring frustration.

2.6 Challenge

Small, easy challenges are great for getting people excited. When challenges are simple, people will try them. With successfully solving challenges by spending time playing, people will enjoy the experience and keep playing to see what comes next. This connects with point 11. Achievement.

2.7 Customize

Every human being has a compelling need to express themselves. When you allow users to use their creativity to create something or add their personal touch, they will invest time, have fun and find the task more valuable. To add a personal touch to those cubes, food coloring or eco-friendly color paint could be provided for sugar and wooden cubes, respectively. Although adding color is not necessary for the challenges, self-expression goes a long way. Coloring the twelve cubes without affecting the challenges should follow the following rule: Use only three different colors to color or paint every four cubes with the same color. Example colors are purple, orange, and green [14].

2.8 Feature

The ability to notice secret features can be an interesting way to motivate people to complete challenges. A preview of these features is shown at the beginning so that players know that there is something to notice. In the exponential function, using twelve cubes is not enough. The tips for using negative space arouse the player's curiosity and add an element of surprise (9. Curiosity and 10. Surprise). Moreover, features such as important statistical concepts are introduced gradually. For example, the probability of each event is always between zero and one (Challenge 3) and the sum of probabilities always equals one (Challenge 8). Certain concepts are repeated for different statistical distributions. For instance, the word discrete is explained for both a discrete distribution and a discrete random variable (Challenges 7 and 8).

2.9 Curiosity

Curiosity is a compelling way to keep players interested. The sequence of challenges sparks curiosity and the desire of players to know how the story ends.

2.10 Surprise

Making sure that the basic elements of each challenge are recognizable (*e.g.*, game flow). This way, players understand and get used to the challenges. They find out how to play and what to expect. New and unexpected elements keep the challenges exciting and add surprise. Challenge 5. Poisson distribution mixes the Discover and Present steps to carry out a Poisson process and then represent its probability distribution. Another aspect that is introduced in later challenges are 'Extra' notes that further explain in statistical terms the challenge.

2.11 Achievement

When users are congratulated on their achievements they will feel proud. This will encourage them to play further. If they don't expect to be encouraged, the effect will be even greater! The third step 'Enjoy' illustrates such an example of persuasion. Progressing through the ten challenges marks each challenge and successfully achieving the solution.

2.12 Competition

By allowing players to compete against each other, an extra layer is added to the game. It is often much more fun to win (or beat someone) than to play alone. Competition encourages players to play more or replay challenges, as they will constantly challenge each other for a rematch. This could also spark discussion and improve their overall understanding.

2.13 Collaboration

Competition is powerful, but so is collaboration. Allowing players to work together and solve challenges is more positive than playing against each other. Players will also learn to communicate better and further discuss the challenges. The different challenges can be played with another player.

3 Challenges

3.1 Challenge 1: Linear function

Discover. Consider each cube a value of 1. From left to right: start with 1 (1 cube), then 2, then 3, then 4.

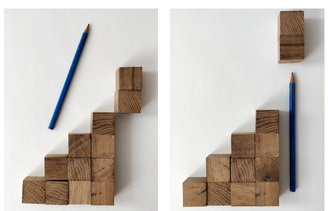
Present. A linear function increases by a constant amount (the value of its slope) in each time interval. Real-world example problems solved by linear functions include age, speed, time and distance, pressure and force.

Enjoy. Congratulations! You just made a linear function that increases by 1. It can be written $f(n) = n + 1$ with n being the number of cubes. You can represent each cube as a square on paper and continue adding 1 square at each iteration. This linear function $f(n)$ increases by 1 (a constant slope) every time n increases by 1.

Figure 2: Linear function

3.2 Challenge 2: Exponential function

Discover. Consider each cube a value of 2. From left to right: start with 2 (1 cube), then 4 (2 cubes), then 6, then 8, then 16. You might wonder now that you are missing some more cubes. Since the value 8 requires 4 cubes, the value 16 will require 8 cubes. To represent this difference, add 4 more cubes to the 4 cubes at value 8. Then, move them to the right. The negative space demonstrates how fast this goes. If you have more cubes you may fill up the negative space or even continue further. Every next value doubles the current one. If you do not have enough cubes, you may draw on paper the space those cubes would occupy or use the pencil as a representation of 6 cubes. What was meant was to contain or limit the uncontrolled exponential growth of infections to prevent hospitals from being overrun.
Present. An exponential function increases by a constant percentage (or ratio) in each time interval. When you recall the beginning of the COVID-19 pandemic, governments were pushing to flatten the curve.
Enjoy. Well done! You just made an exponential function that increases by 2^n with n varying from 1 to 4. This can be written 2^1, 2^2, 2^3, 2^4 which corresponds to 2, 4, 8, and 16, respectively. It can be written $f(n) = 2^n$. This exponential function $f(n)$ increases by 50 % (a constant percentage) every time n increases by 1.



(a) (b)

Figure 3: Exponential function

3.3 Challenge 3: Uniform distribution

Discover. This challenge requires all the cubes in a bag. You are also welcome to ask a friend to join you for this challenge. Draw one cube and declare it your favorite. Set it aside for a second, then put it back with the others. Place them all inside the bag, then without looking draw your favorite cube. Depending on the kind of cubes you use, it may be simply unidentifiable or impossible to tell which is which. It's completely normal, this is a difficult task. After drawing one cube, put it back in the bag. In the best case scenario, you will draw your own cube on the first try. In the worst case scenario, your favorite will never be drawn.
Present. A uniform distribution or a rectangular distribution is a distribution that is concerned with events that are equally likely to occur. It has constant probability. Assuming there's no preference for any particular cube, you'd imagine that the probability of each of the twelve cubes $1, 2, 3, \ldots, 12$ is the same. Since all the probabilities must add up to 1, a logical conclusion is to assign a probability of $1/12$ to each of the 12 options. A deck of cards also has a uniform distribution. This is because a person has an equal chance of drawing a spade, heart, club, or diamond. Another example of a uniform distribution is a coin toss. The probability of getting a tail or a head is the same.
Enjoy. Kudos! You just experienced the uniform distribution through a randomized experiment in which a drawn item is always returned to the universe of possibilities before the next draw. The probability that you draw your favorite cube follows a uniform distribution because every cube is equally likely to be drawn next. We can write the probability P of drawing any cube x from 1 to 12 as $P(X = x) = 1/12$ for $x = 1, 2, 3, \ldots, 12$. In statistical terms, we say that the probability of the drawing event remains the same throughout the experiment. To understand a uniform distribution, we have to capture the probability of X being close to a single number. This is achieved by relying on a probability distribution function. When the probability distribution around a point x is large, it means the random variable X is likely to be close to x. If the probability distribution around a point x equals zero, it means that X won't be in that interval. The probability of each event is always between 0 and 1.

Figure 4: Uniform distribution

3.4 Challenge 4: Normal or Gaussian distribution

Discover. Consider each cube a value of 2. From left to right: start with 2 (1 cube), then 4, then 6. You might wonder now, this looks much like a linear function. Continue adding cubes but in reverse by restarting from the end. Add again 6, then 4, then 2. The second part looks very much like an inverted version of the first part. It is as if a mirror is placed in the middle. You may use your pencil as if it was a mirror.
Present. A normal distribution is a probability distribution where the values are concentrated in the middle of the interval and the rest shrink symmetrically toward the two extremes. While statisticians and mathematicians uniformly use the term 'normal distribution', physicists sometimes refer to it as a Gaussian distribution and, because of its curved shape, social scientists call it the 'bell curve'. It represents real-valued random variables whose distributions are not known to us: the average height of NBA players, the average female shoe size, the volume of milk production from cows, the shifting distribution of summer temperatures, *etc*.

(a) (b)

Figure 5: Normal or Gaussian distribution

Enjoy. Well done! You just discovered the normal distribution. A normal distribution in a variate X with mean μ and variance σ^2 is probably the most important statistical distribution to know because many phenomena correspond to this distribution (*e.g.*, height, shoe size). The normal distribution is symmetric around the mean, showing that data close to the mean are more frequent than data far from the mean. A typical normal distribution looks much like the one you just built. It has a mean $\mu = 0$ and a standard deviation $\sigma = \sqrt{\sigma^2} = 1$.

The complete ten challenges are available as supplementary.

3.5 Challenge 5: Poisson distribution

3.6 Challenge 6: Exponential distribution

3.7 Challenge 7: Gamma distribution

3.8 Challenge 8: Bernoulli distribution

3.9 Challenge 9: Binomial distribution

3.10 Challenge 10: Chi-squared distribution

3.11 Extra Challenge: Distribution Comparisons

4 Discussion

This work is designed as a proposal to begin discussing ways to improve scientific communication and public understanding of basic statistical concepts. The work is limited by the number of cubes used. While the examples are dynamic physicalization examples, using more cubes or even LEGOs would help players record, update, and explore the differences between the 10 challenges. This work is inspired by data-driven presentations such as Hans Rosling's explanation of global population changes. Indeed, this work borrows the concept of unit elements, much like Isotype visualizations and to build visualizations. While this has not yet been tested, there is considerable room for improvement in the presentation and potentially in the real-world examples offered. User-based evaluation would provide details about the needs and problems encountered by a player and often results in a set of suggestions for improving the player experience for different user groups. The 10 challenges and explainable analogs of growth functions and distributions proposed for statistical literacy and fluency follow thirteen initiatives, including ten gamification techniques and three data materialization motivations. We adopt the latter as steps for gamification design. The first, "Discover" has an inherent surprise effect that is offered by the progressive orientation of building a visualization. Although it is often not clearly stated what the objective is. It builds the statistical concept implicitly. Although the player is guided through the construction process, low-level details are completely absent and high-level understanding is not yet possible. The goal is to discover the instinct behind the concept that we want to present, explain, appreciate and consume later. The second part, "Present" explains what the player(s) have built and highlights concrete examples and uses of the underlying statistical concept. This tug-of-war between simplicity and complexity adds to the element of surprise and makes the complex concepts more accessible and ultimately more digestible. Although the goal is to present relatable examples, some examples may be more relatable to people in one part of the world than to those in another. Explainability is affected by cultural interpretations or considerations. As with other games, it would be desirable to create region-specific examples (*e.g.*, Cards Against Humanity editions: Canada, UK, International, *etc*). The third and final section, "Enjoy" presents domain-specific notations and statistical jargon after congratulating the player(s) for completing a challenge. Successful completion of the challenge is important and provides a level of satisfaction that can be used to learn new things. The consumption of new information also follows a progression of difficulty and repetition of previous challenges. While this can be successful, it is not clear what the player(s) can understand without conducting a user-based assessment. Additionally, this work reveals a link to visualization literacy [15]. The assignment of a value to a cube serves as a parallel to an encoding channel through which data is translated into visualization. Further parallels could be drawn from a proposed framework evaluating data visualization literacy [16]. The 10 challenges are sequential. In other words, the order of the challenges is important and must be followed. This is done by design, as a story, the important elements for understanding the later challenges are introduced in the previous challenges. Like breadcrumbs in a story, additional elements are sometimes added after players have completed a challenge.

Acknowledgments

Finding a new perspective or discussing concepts with fresh eyes is always helpful. Many thanks to Aleksandar Anžel, Chisom Ezekannagha, and Theresa-Marie Rhyne for their feedback and encouragement.

References

[1] Elnaz Safapour, Sharareh Kermanshachi, and Piyush Taneja. A review of nontraditional teaching methods: Flipped classroom, gamification, case study, self-learning, and social media. *Education Sciences*, 9(4):273, 2019.

[2] Eser Çeker and Fezile Özdaml. What" gamification" is and what it's not. *European Journal of Contemporary Education*, 6(2):221–228, 2017.

[3] Cigdem Hursen and Cizem Bas. Use of gamification applications in science education. *International Journal of emerging technologies in Learning*, 14(1), 2019.

[4] Trevor Hogan, Uta Hinrichs, Yvonne Jansen, Samuel Huron, Pauline Gourlet, Eva Hornecker, and Bettina Nissen. Pedagogy & physicalization: Designing learning activities around physical data representations. In *Proceedings of the 2017 ACM conference companion publication on designing interactive systems*, pages 345–347, 2017.

[5] Pierre Dragicevic, Yvonne Jansen, and Andrew Vande Moere. Data physicalization. *Handbook of Human Computer Interaction*, pages 1–51, 2020.

[6] Alejandro Canales, Alejandro Pena, Ruben Peredo, Humberto Sossa, and Agustin Gutierrez. Adaptive and intelligent web based education system: Towards an integral architecture and framework. *Expert systems with applications*, 33(4):1076–1089, 2007.

[7] Mei-Hua Hsu. A personalized english learning recommender system for esl students. *Expert Systems with Applications*, 34(1):683–688, 2008.

[8] Mu-Jung Huang, Hwa-Shan Huang, and Mu-Yen Chen. Constructing a personalized e-learning system based on genetic algorithm and case-based reasoning approach. *Expert Systems with applications*, 33(3):551–564, 2007.

[9] Martin Schonger and Daniela Sele. How to better communicate the exponential growth of infectious diseases. *PLoS One*, 15(12):e0242839, 2020.

[10] Joris Lammers, Jan Crusius, and Anne Gast. Correcting misperceptions of exponential coronavirus growth increases support for social distancing. *Proceedings of the National Academy of Sciences*, 117(28):16264–16266, 2020.

[11] Samuel Huron, Yvonne Jansen, and Sheelagh Carpendale. Constructing visual representations: Investigating the use of tangible tokens. *IEEE transactions on visualization and computer graphics*, 20(12):2102–2111, 2014.

[12] Katie Seaborn and Deborah I Fels. Gamification in theory and action: A survey. *International Journal of human-computer studies*, 74:14–31, 2015.

[13] Lila A Loos and Martha E Crosby. Gamification methods in higher education. In *International Conference on Learning and Collaboration Technologies*, pages 474–486. Springer, 2017.

[14] Georges Hattab, Theresa-Marie Rhyne, and Dominik Heider. Ten simple rules to colorize biological data visualization, 2020.

[15] Elif E Firat, Alark Joshi, and Robert S Laramee. Interactive visualization literacy: The state-of-the-art. *Information Visualization*, 21(3):285–310, 2022.

[16] Katy Börner, Andreas Bueckle, and Michael Ginda. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences*, 116(6):1857–1864, 2019.