Linking Probability and Statistics at the K-12 Level

A Brief produced at the Park City International Seminar
Park City Mathematics Institute July 3-8, 2017

> Both statistical reasoning and probabilistic reasoning alone have their limitations and their merits. Their full power for advancing human knowledge comes to bear only in the synthesis acknowledging that they are two sides of the same coin.
>
> Batanero, et al. (2016)

Working Definitions of Probability and Statistics

For the purposes of this brief, probability is referred to as the study of chance (and the likelihood/certainty that an event will happen or not). Statistics will be considered to be the study of collecting, analyzing, and interpreting data. From these perspectives, the role of probability with respect to statistics envisioned agrees with Batanero, et al. (2016, p. 11) that "Indeed, it is here where data and chance (i.e., statistics as the science of analyzing data and probability as the study of random phenomena) come together to build the powerful foundation of statistical inference."

There are, of course, highly mathematical applications of probability. However, many of those applications are usually not encountered at the pre-university level.

Two Stages of Education: Elementary (K-5 or K-6) and Secondary (up to Pre-University)

Uses of technology and specific activities that may help forge links between probability and statistics will be presented below for each of the two stages of education. There are wide differences in what countries include in their formal school curricula. However, with respect to probability the activities presented below the following curricular elements at each level are assumed:

- Elementary: informal introduction to concepts of likelihood and certainty through words/descriptors, and
- Secondary: formal probability including investigations of chance processes and probability models to eventually include random processes, conditional probability and the use of probability to aid in decision making.

In some countries, such as the United States, probability in high school may be treated only as a tool for statistics (Franklin et al., 2007, p. 8), and concepts such as conditional probability may only be dealt with informally. In other countries, such as the Philippines and the Czech Republic, the high school curriculum might call for a quite formal treatment of probability concepts.

The personal progress of each student should not be forgotten. The development of a deep understanding is a lifelong process for each individual. Education in probability and statistics should ensure sense-making and critical thinking in various contexts (i.e. cultural, social, biological, economical, etc.).

Providing Conceptual Links between Probability and Statistics

An important link between probability and statistics is analyzing and interpreting data to make predictions. In other words, students need to learn how probability links to the collection and analysis of data to make decisions under uncertainty.

Hurvich (2010) suggests that the link between probability and statistics can be succinctly characterized as

> With statistics, you go from observed data to generalizations about how the world works. For example, if we observe that the seven hottest years on record occurred in the most recent decade, we may conclude (perhaps without justification) that there is global warming. With probability, you start from an assumption about how the world works, and then figure out what kinds of data you are likely to see [i.e., what predictions can be made from the data].

In K-12 descriptive statistics, students are exposed to the processes involving data collection, representation and interpretation. Specifically, for numerical data, several measures can be used to describe a data set, e.g. center, location, variation, skewness, etc. Bakker and Gravemeijer (2004) posit that the treatment of data on numerical measures leads to ideas about population distributions. Given distributions for which probabilistic qualities are known, it may be possible to make predictions for individual data. Meanwhile, progressing towards continuous probability distributions exposes the students to the ideas of not only representing data plots as a curve in the plane but also that of computing probability as the area of the region beneath this curve.

Topics from inferential statistics such as hypothesis testing are not in the official high school curriculum of all countries. However, they do appear in many and can show the power of linking probability and statistics with considerations of sampling, randomness, levels of significance related to a *p*-value for a region of acceptance or non-acceptance of a null hypothesis, confidence intervals, sampling distributions, etc.

As students begin to work with samples, they may encounter the concept of *sample bias* (the situation in which all elements of the population do not have an equal chance of being in the sample). It is challenging to understand that even with random assignment of samples, variation can be expected from sample to sample, but "when randomness is incorporated into the sampling procedure, probability provides a way to describe the 'long-run' behavior of this sampling variability" (Franklin et al., 2007, p. 66).

Understanding and confronting sample bias may be particularly challenging in cultural contexts where there may not be a local natural language or academic language equivalent of the word *bias* or its use in the term *sample bias*. In English, the popular meaning of the word *bias* has to do with prejudice, or having already made up your mind about something and being unwilling to consider facts that conflict with your point of view. Even in cultures that have word(s) for such bias, there may not be an academic term for *sample bias* and something like a direct translation of *systematical error* might be used (as in Czech). Some cultures may not traditionally use the concept or have language for such terms as *chance* and *future*. They may borrow terms from a national language, but students may face an extra challenge if those concepts are not natural to them.

Technology

Technology has impacted the teaching and learning experiences inside and outside the classroom — as echoed by the National Council of Teachers of Mathematics (2000, p.3): "Technology is essential in teaching and learning mathematics; it influences the mathematics that is taught and enhances students' learning."

Nowadays, technology centers on electronic digital devices, encompassing several media to deliver information. Computer technology has significantly progressed with hardware and software that can enhance the computing experience in schools. New developments, even cell phone applets, allow users to quickly compute probabilities based on vast amounts of data, simulate complicated probabilistic situations and create technology-based learning situations to explain probabilistic concepts.

In mathematics education, classrooms and laboratories equipped with computers, digital projectors, and/or interactive whiteboards can be used by both teachers and students for demonstration, collaboration, and interactive activities. Moreover, handheld devices such as smartphones and tablets give education users not only mobility, but also access to mathematics applications that can enrich the experience. Additionally, scientific and graphing calculators have constantly been updated to include new features that allow richer and more varied experiences with technology. Mathematics software is available that can be used not only in computations but also in simulations — an important role of technology in probability education. Examples include spreadsheets, statistics software, modeling software, and specialized applications and applets that can generate random numbers or simulate a coin toss, dice roll, etc. Such applications can very quickly present repeated simulations that can aid in a conceptual understanding of randomness.

It should be pointed out that hands-on activities with manipulatives such as coins, dice, spinners, colored-objects in bags, etc. should come before using electronic technology that has the advantage of very rapidly simulating many trials.

Sample Activities at the Elementary and Secondary Levels

Students often have very different abilities and personal dispositions. The activities below should allow students with a wide variety of mathematical backgrounds and dispositions to gather, process and evaluate data and use these data to answer probability questions. While working on these activities, discussions between the students and between students and teacher are important for understanding concepts if the activities are to lead to a comprehensive understanding of probability.

The following activities should contribute to an understanding of the link between probability and statistics, and to the development of the critical thinking.

Elementary Level

Task 1: A survey on categorical data

Goal: These informal activities involve students in gathering data, responding to questions on probability through words pertaining to likelihood, and then extending the data collection by interviewing students in other classroom to see how predictions from their first data set are verified as they extend the data collection.

Introduction: During the summer holidays the students traveled with their parents or with their friends to some destination or they stayed at home.

Question: Where did you spend your holidays?

It should be noted that the question can be changed to another that is relatable to the students.

Activity 1: Each student records where he/she was during the school holiday. He/she chooses only one destination. The teacher reports the results.

Data collection: The students should decide how to divide the data into groups, for example by the country in which the destinations are located, or by the location (e.g. by the sea, in the mountains, in the cities, etc.).

Data processing: The students organize the data into a table. They distribute the places as required and indicate the frequencies in each country, region or location.

Evaluation of data: Discussion about the data collected in class.

Appropriate questions:

- Which destination is most/least likely to be visited?
- If you go to another class at the same or similar level and ask any one student the same question, what do you think he/she will answer?
- What is the likelihood they will go to _____?

Activity 2: Each student interviews another student from another class during their free time.

The first question they should ask is “Has my classmate asked you about where you spent your holidays?”. If the answer is “yes”, the student should find another to ask. Otherwise, he/she asks “Where did you spend your holidays? Choose only one destination.”

Evaluation of data: Discussion about the data collected in another class.

Appropriate questions:

- In the other class, which destination was most/least likely visited?
- How do the results compare to those from our class?
- If we combine the information from both classes, which destination will most/least likely be visited?

Concluding steps: Discussion and comparison of new data with past outcomes.

Optional: The students ask more students (for example: 20 respondents) or students from other grade levels. The questions are the same as in Activity 2.

Computer tools: Projector or Interactive Whiteboard, spreadsheet software (e.g. Google Spreadsheet), cloud storage (e.g. Google Docs)

Data collection: Each student inputs his/her data into the table using Google Docs or something similar.

Data processing: Students use the technology to sort and organize the responses

Evaluation of data: Discussion about the data collected in the school.

Appropriate questions: The same as in the activities 1 and 2.

Task 2: Weather forecasts and necessary preparations

Goal: This informal activity involves students gathering data and making necessary preparations. It also highlights quantitative literacy.

Introduction: Students record the weather each day in their personal journals. Before recording the data, they should, as a class, decide what to record and what words to use so the data are somewhat alike.

Question: What is the chance of rain in the upcoming days?

Activity: Students are tasked to check the chance of rain for the next three days. They may use the following to gather necessary data, among others: a reliable weather app in their smartphone, a trusted search engine, news in the television or newspaper. On the following day, they will share their findings.

It should be noted that chance of rain pertains to the product of “the confidence that precipitation will occur somewhere in the forecast area” and “the percent of the area that will receive measurable precipitation, if it occurs at all.” (National Weather Service, 2017) It is a misconception that chance of rain correlates to the strength of the precipitation.

As an example: Students find out that there is an 80% chance of rain all throughout the next three days.

Optional: Ask the students to find data on precipitation for the last 50 years for their region. Open a discussion on future weather patterns and the possible impact on society. Allow students to reflect on ways to mitigate the negative impacts, if there are any.

Task 3: The random path

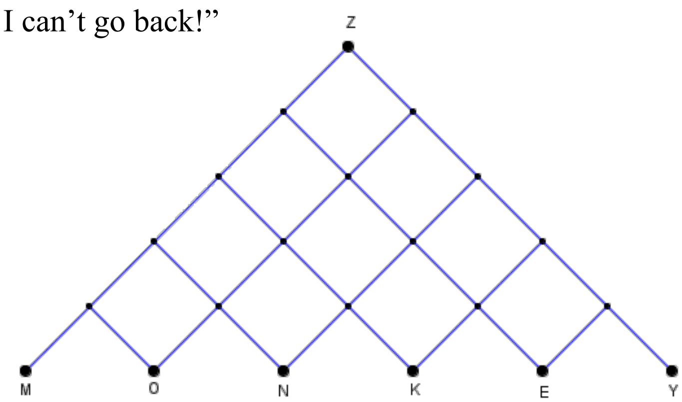
Goal: This task begins the transition from informal to formal presentations of probability concepts.

Introduction: Imagine you are in a town that has perpendicular paths as shown in the diagram below. Between the paths, there is a lawn (the squares) that cannot be walked on.

At the end of the paths, there are baskets with different candies – maroon, orange, navy blue, khaki, emerald green and yellow (in the picture: points M, O, N, K, E, Y).

Every type of candy is delicious, and you cannot decide which one you want.

You stand in the uppermost corner (point Z) and you leave it to chance. In your pocket, there is one coin. You decide: "At every crossroad, I throw a coin. If it will be a head, I'll go left; if it will be tail, I'll go right. I can't go back!"



Question: Do you think you have equal chances of getting each candy?

The teacher can prepare a copy of the figure for each student.

Activity 1: The students throw the coin five times and reach a candy.

Data collection: Each student reports to the teacher which candy was received. The teacher places the results in a table.

Appropriate question:

- Which candy is most/least likely?

Conclusion: The number of outcomes is not enough. We have to do more trials.

Activity 2: The students throw the coin five times, get a candy, and repeat the experiment again and again for 10 times.

Data collection: Each student reports to the teacher the candies they reached. The teacher places their results in the table.

Appropriate questions:

- Now that we have more results, which candy is most/least likely?
- Describe the chances of getting each candy.

Evaluation of data: Discussion about meaning of the outcomes.

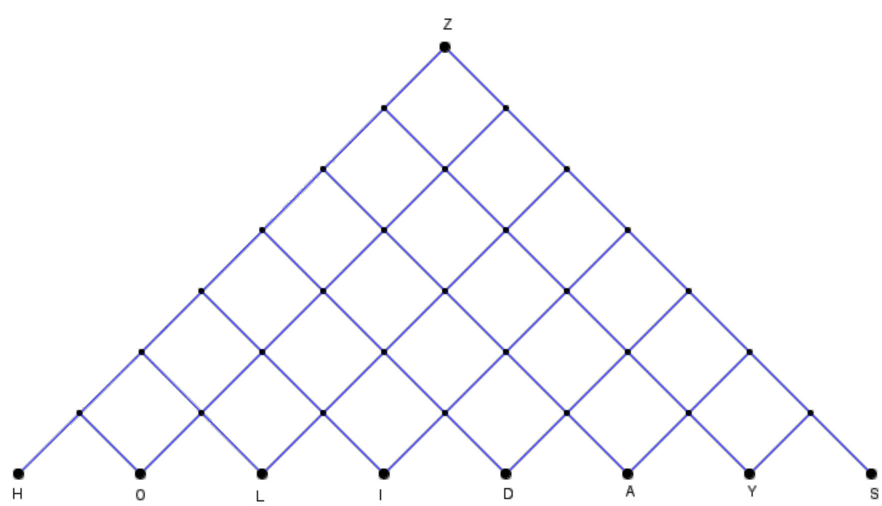
Activity 3: The teacher asks the students to draw a bar graph to show the frequency of the candies for the whole class.

Appropriate questions:

- After seeing the bar graph, which candy has the lowest bar? highest bar?
- What do the bars mean?
- If someone else does the experiment, which candy will be least likely? Most likely?

Conclusion: Connecting to the ideas of probability and binomial distribution: candies N and K are the most likely to be reached, while M and Y are least likely.

It should be noted that the teacher can modify the picture and add more paths (see sample below). For higher grade levels, the teacher can begin posing more formal probability questions utilizing the same figure.



Secondary Level

Task 1: The two-way table and probabilities

Goal: This informal activity tasks students to gather data and compute probabilities, including conditional probability, without using the formula.

Introduction: Children oftentimes aspire to be scientists when they grow up. Given the choice between becoming a doctor or a teacher, which will they choose to be.

It should be noted that the context can be changed to another that is relatable to the students.

Question: Would elementary students rather be a doctor or a teacher when they grow up?

Activity 1: The secondary-level students are assigned to visit an elementary level classroom (either Grades 1-3 or Grades 4-6). They will ask the students in their assigned class "Which would you want to be if you had the chance: a teacher or a doctor?"

Technology (recommendation): Projector or Interactive Whiteboard, spreadsheet software

Data collection: The students record their own collected data.

Data processing: The students combine the data into one two-way table (see sample below) either by hand or in a spreadsheet.

	Grades 1-3	**Grades 4-6**	**Total**
Teacher			
Doctor			
Total			

Evaluation of data: The teacher and the students discuss the data and observe the differences between the two age groups.

Appropriate questions:

If you ask a random student in the elementary level about a career preference, what is the probability that the person

- wants to be a teacher? a doctor?
- is from Grades 1-3? Grades 4-6?
- wants to be a doctor, given that the person is from Grades 4-6?

- is from Grades 1-3, given that the person wants to be a teacher?

Task 2: Hands-on coin tossing

Goal: This formal activity involves students collecting data, then comparing the theoretical and experimental probabilities they obtain.

Introduction: Students have learned that the theoretical probability of getting either a head or tail after flipping a fair coin is 0.5. This activity might pose a conflict if the results favor either heads or tails. The aim is to emphasize that after a large number of trials, the probability of getting a head or tail approaches being equally likely[1].

Question: After recording and combining the results of each student's coin tosses, how likely is it to get a head or a tail?

For this activity, all students should use the same kind of coin. It is assumed that the coins are fair. Teachers may choose to preempt students with the idea that the way the coins are flipped and the material used in the coins, among other factors, may introduce bias in the experiment. As an example, assume there are 30 students in the class.

Activity 1: The activity is divided into three trials: each student does the coin toss for 20, 30, and 50 times. In each stage, they record their results on the board or report them to the class secretary. To enable them to experience the variability among the samples, students should each compile the data on their own, comparing answers to be sure they are correct. .

Appropriate questions:

- First trial: After tossing the coin for a total of 600 times, what is the chance of getting a head? a tail?
- Second trial: After tossing the coin for a total of 900 times, what is the chance of getting a head? a tail? How different are the results now in comparison to the first?
- Third trial: After 1500 tosses for the entire class, what is the chance of getting a head? a tail? What does the activity tell us about experiments with few trials, then increasing the number of trials?

Even though the probability of an event is 0.5, in the experimental probability, there will be variability, and, if the coin is fair, in the long run will approach 0.5.

Activity 2: Using appropriate software such as Texas Instrument 83 or 84 calculators, Nspire or any spreadsheet software (e.g. Microsoft Excel, Apple Numbers), simulate the coin toss with the number of trials becoming larger and larger.

[1] After doing this activity, it might be interesting for the students to know that the British mathematician John Kerrich, while in a German prison camp in World War II, along with a fellow prisoner tossed a coin 10,000 times and got 5,067 heads (https://en.wikipedia.org/wiki/John_Edmund_Kerrich).

Optional: Pose the problem "Suppose a student tosses a coin a hundred times, and gets 40 heads. Is the coin fair? Why?" How about 25 heads?

Task 3: A survey on numerical data

Goal: This task involves students gathering data outside the school and making interpretations for a non-random sample, thus extending the development of informal inference.

Introduction: What do we know about the inhabitants of our town?

Question (choose just one):

1. How old are they? (chosen for this sample activity)
2. How tall are they?
3. How many children do they have?

It should be noted that the question can be changed to one that is relatable to the students.

Activity: The students in class are grouped such that each group will be assigned to a particular street to avoid repetition. During their free time, the groups will conduct a survey on the age of the inhabitants in the assigned street. The goal is to have at least 100 respondents.

Computer tools: Computer laboratory, projector or Interactive Whiteboard, spreadsheet software

Data processing: In class, the students discuss the data that the whole class collected and decide on appropriate intervals for entering the data on a spreadsheet. Each student group:

- computes the relative frequency for each interval (expressed as a percent)
- represents the data graphically
- finds the mean, median and measures of spread.

Afterwards, the group discusses the advantages of using a histogram or dot plot to graphically represent the data.

Appropriate questions:

- What is the average age of the inhabitants in our locality?
- What is the most common age bracket of the inhabitants?
- When you ask a random person in our locality about their age,
  - (a) What is the most likely age?
  - (b) What is the likelihood that a person's age belongs to the interval _____?
- Comparing with government data, are our results credible?

Culminating steps: The students summarize their research by making interpretations of the data they gathered. The teacher connects the results to ideas of probability and should mention that the sample size is very important for the credibility of the research.

Optional: Suppose that random sampling was used in conducting this survey. What do the results tell about the population of the town?

.

Conclusion

Data sets are readily available from reliable sources on the Internet (e.g. government websites) or institutions. In light of these data, how do we prepare for the future?

Students can be grouped to study a simple research topic that concerns the community (e.g. weather, economy, agriculture, etc.) A question has to be formulated and addressed at the end of the activity. The groups may obtain their data from reliable websites such as the government, academia, or other private institutions. Using the appropriate statistical treatment, each group conducts hypothesis testing and provides an interpretation of results.

Some questions that can be considered:

- What do the results mean for our community? (e.g. Will the food supply be adequate?)
- If the results are positive, how can we sustain them?
- If the results are negative, what can we do to improve the situation?

It is worth noting that the questions may change depending on the contexts involved in the research done by the students. Regardless, questions should revolve on the uncertain future.

A robust education in the use of probability can lead to informed citizens. Data are everywhere, but they must be validated. Then notions of probability can be used to work with verified data to make wise decisions. Probability questions can be asked about data that has been collected and analyzed, and, at the inferential level, probability must be used.

References

Bakker, A., & Gravemeijer, K.P.E. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds). *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147-168). Dordrecht, The Netherlands: Kluwer Academic Publishers

Batanero, C., Chernoff, E., Engel, J., Lee, H. & Sánchez, E. (2016). *Research on teaching and learning probability*, ICME-13 Topical Surveys, DOI 10.1007/978-3-319-31625-3_1

Franklin, C., Gary Kader, G., Mewborn, D., Moreno, J., Peck, P., Perry, M. & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education* (GAISE) *report: A Pre-K –12 curriculum framework.* Alexandria, VA: American Statistical Association.

Hurvich, C. (2010) *Probability theory.* Retrieved from http://pages.stern.nyu.edu/~churvich/Undergrad/Handouts1/07-Prob.pdf.

National Council of Teachers of Mathematics (2000). *Executive summary of principles and standards for school mathematics*. Reston, VA: 2000.

National Weather Service. (2017). Retrieved from https://www.weather.gov/ffc/pop

Contributing Authors

Dvořáková, Blanka; Gymnázium Teplice, Teplice, Czech Republic
Giménez, Joaquin; Universidad de Barcelona, Barcelona, Spain
Guzon, Angela Fatima H.; Ateneo de Manila University, Quezon City, Philippines
Hao, Lester C.; Chiang Kai Shek College, Manila, Philippines
Inekwe, Israel; Michael Okpara University of Agriculture, Umudike, Nigeria
Mejía, Bayardo; Universidad de San Carlos de Guatemala (USAC), Guatemala City, Guatemala
Sánchez, Mario; The American School of Guatemala, Guatemala City, Guatemala
Scott, Patrick; Inter-American Committee on Math Education (IACME), Santa Fe, NM, USA
Serradó Bayés, Ana; Colegio La Salle Buen Consejo, Cádiz, Spain
Spěvák, Jan; J.E. Purkyne University, Ustí nad Labem, Czech Republic
Teague, Daniel; The North Carolina School of Science and Math (NCSSM), Durham, NC, USA

Reviewers

Burrill, Gail; Michigan State University, East Lansing, MI, USA
Friedberg, Solomon; Boston College, Chestnut Hill, MA, USA