Student Estimates of Probability and Uncertainty in Advanced Laboratory and Statistical Physics Courses

Donald B. Mountcastle[1], Brandon R. Bucy[1], and John R. Thompson[1,2]

[1]*Department of Physics and Astronomy and* [2]*Center for Science and Mathematics Education Research*
University of Maine, Orono, ME

Abstract. Equilibrium properties of macroscopic systems are highly predictable as n, the number of particles approaches and exceeds Avogadro's number; theories of statistical physics depend on these results. Typical pedagogical devices used in statistical physics textbooks to introduce entropy (S) and multiplicity (ω) (where $S = k \ln(\omega)$) include flipping coins and/or other equivalent binary events, repeated n times. Prior to instruction, our statistical mechanics students usually gave reasonable answers about the probabilities, but not the relative uncertainties, of the predicted outcomes of such events. However, they reliably predicted that the uncertainty in a measured continuous quantity (e.g., the amount of rainfall) does decrease as the number of measurements increases. Typical textbook presentations assume that students understand that the relative uncertainty of binary outcomes will similarly decrease as the number of events increases. This is at odds with our findings, even though most of our students had previously completed mathematics courses in statistics, as well as an advanced electronics laboratory course that included statistical analysis of distributions of dart scores as n increased.

Keywords: Thermal physics, statistical mechanics, statistics, multiplicity, entropy, Second Law of Thermodynamics, equilibrium properties, probability, uncertainty, coin flips, physics education research
PACS: 01.40.-d, 01.40.Fk, 05.70.Ce, 65.40.De, 65.40.Gr.

INTRODUCTION

At the University of Maine (UMaine), we are actively involved in a research study of the teaching and learning of thermal physics at the advanced undergraduate level [1,2]. One theme of our research is the extent to which student mathematical conceptual difficulties may affect understanding of physics concepts. Our department offers both a one-semester classical thermodynamics course (text: Carter [3]) and a separate one-semester course in statistical mechanics (text: Baierlein [4]). Here we report research on student understanding of probability and statistics in three successive semesters of *Statistical Mechanics* (Spring 2005-07), along with a statistics project in a junior level *Physical Electronics Laboratory*.

Previous work in PER has explored introductory student understanding of measurement error analysis and measurement uncertainty [5-7]. We are investigating the extent to which advanced students connect measurement uncertainty in a laboratory context to probability and uncertainty in statistical physics.

Twenty-seven *Statistical Mechanics* students are included in this data; most were senior physics majors, two were mathematics majors and four were physics graduate students. Seventeen of the twenty-three undergraduates in *Statistical Mechanics* included in this study had previously participated in the *Physical Electronics Laboratory* course, either the previous semester or a full year earlier; sixteen had previously taken a mathematics course in statistics.

The theoretical foundation of the learning objectives reported here is the *fundamental postulate in statistical mechanics*: All accessible microstates are equally probable in an isolated equilibrium system. Mathematical foundations are the *Central Limit Theorem*, and the *Strong Law of Large Numbers*.

Boltzmann gave us the extraordinary insight that the world is highly predictable *due entirely to probability*, expressed in multiplicity *(ω)*, entropy *(S)* (where $S = k \ln(\omega)$), and the Second Law of Thermodynamics. Statistical mechanics textbooks [3,4,8] typically introduce the definitions and meanings of multiplicity and entropy by exploring the binomial coefficient describing independent binary outcomes of n events (such as coin flipping) as n increases. The advantage of this approach is that it offers concrete countable events that can be easily visualized, measured, calculated and displayed as histograms of probability, shown to converge, *with overwhelming probability*, to a single predictable system macrostate; i.e., the inevitable equilibrium state. Convergence is seen as a transition from a discrete, broad histogram (small n) into a continuous, smooth, sharply-peaked function (large n), accompanied by a decreasing relative uncertainty ($\Delta n/n$) (Figure 1). Full understanding of that transition requires the conceptual ability to move seamlessly from a summation to an integral, from obvious to hidden degeneracy, and from discrete states to a (continuous) density of states. This conceptual ability is fundamentally important in learning statistical mechanics [9], yet at the same time can be exceptionally challenging for students.

Often textbooks thus use independent binary outcome events, but rather quickly extrapolate from small n to large n, emphasizing the narrowing sharpness of the macrostate (probability peak; Figure 1) as n increases toward the thermodynamic limit, all based on simple rules of combinations, and probability. The (vanishing) *uncertainty* associated with the most-probable macrostate is usually acknowledged only indirectly by phrases such as *overwhelming probability,* or *narrowing peak-width*.

An exercise during the first week of our *Physical Electronics Laboratory* is a class group project, labeled the "darts project," adapted from Squires (2nd edition) [10]. Each student throws a specified number of darts at the center of a one-dimensional target, reports their scores (based on horizontal location) to the class electronic conference, analyzes and graphs the entire data set ($n > 200$), and compares this result to a Gaussian distribution.

One of the learning objectives of the laboratory course is facility with measurement error analysis and propagation of errors. Thus, the primary reason for the statistics of the darts project is to familiarize the students with an easily visualized, actual numerical distribution that can be seen to grow into a near Gaussian distribution as n, the number of reported scores, increases from ~ 20 to > 200. Students' observations of how the standard deviation (σ) and standard deviation of the mean (σ_m) of the dart score distribution change (or not) with increasing n are relevant to our questions about uncertainty reported here.

QUESTIONS, RESULTS, AND ITERATIONS

We were interested in whether or not our *Statistical Mechanics* students had the inclination to follow and to adopt the typical textbook arguments, using binary event outcomes, that a *single* equilibrium macrostate, *with vanishing uncertainty,* will emerge as n increases towards the thermodynamic limit. These arguments articulate the microscopic basis for maximum multiplicity, entropy and the Second Law. We designed the diagnostic question Q1 (Figure 2) to probe our students' reasoning in this context both pre- and post-instruction.

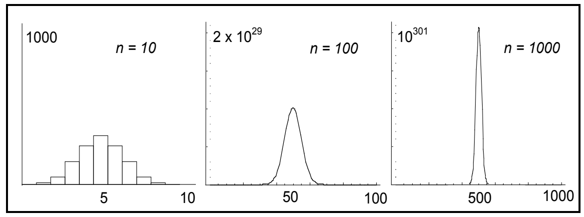

FIGURE 1. Typical textbook introduction to multiplicity, showing the convergence toward $n/2$ of the distribution of n independent binary events from a discrete histogram to a continuous smooth function for large n.

Estimate *all* of your answers as $x = \underline{a} \pm \underline{\Delta a}$. That is, give your best numerical *answers* for both (a) ***and*** numerical estimates of the *uncertainties* (the $\pm \Delta a$) that indicate your confidence level in the precision of your answers (*be realistic*).

If you flip one coin n times, estimate the value of and the uncertainty of the number of 'heads' that would occur for:

a) $n = 4$
b) $n = 100$
c) $n = 1000$
d) $n = N_A = 6.022 \times 10^{23}$

Explain the reasoning for your estimates in (a) - (d) above.

FIGURE 2. Question Q1, given in *Statistical Mechanics* both pre- and post-instruction about multiplicity and probability (pretest version shown).

It is important to point out that we are looking for the qualitative trend of *relative* uncertainties ($\Delta n/n$), as n increases, rather than precise numerical answers. Qualitatively, a prediction of $n/2 \pm n/2$ is *not unreasonable* for $n = 4$ coin flips, but is dubious for $n = 10$, extremely unreasonable for $n = 100$, and totally unreasonable for $n = 1000$ and $n = N_A$. On the pretest, essentially all students as expected correctly gave $n/2$ as their predicted cumulative number of heads. However, their responses for the "realistic uncertainty" were surprising. Fewer than half of the students seemed to believe with confidence in a decreasing trend in relative uncertainty ($\Delta n/n$). A substantial number ($\sim 20\%$) specifically stated uncertainties of $\pm n/2$, for parts (a) – (d), as if all n (or zero) heads is a realistic possibility (Table 1). Although not as severely unreasonable, another 13% predicted the same relative uncertainty (e.g., $\Delta n/n$ of 15%) in outcome for n=100 and n=1000. (Often they gave smaller relative uncertainties for part (d).) Another 17% refused to commit to any model for the n-dependence of uncertainty by omitting it in their responses to Q1. Even those students who did exemplary work previously on the darts project did not always estimate a decreasing uncertainty with increasing n, which was one of the intended learning outcomes of that exercise. On similar post-test questions about coin flips as functions of n (2006-07, N=12), essentially all students gave realistic estimates for both the values *and* uncertainties for the number of heads, although generally with superficial explanations of their reasoning (Table 1).

TABLE 1. Results from Q1 for relative uncertainties.

Response	Pre-instruction (N=24)	Post-instruction (N=12)
Decreases with n (correct)	11	11
Covers entire range	5	--
Same for n=100, n=1000	3	--
No uncertainty	4	--
Other	--	1

From the laboratory class and elsewhere, we know that students quite often accept, even take for granted, that taking multiple readings and averaging those values reduces uncertainties in measured quantities. So, in class the next two years we again gave students Q1, collected their written answers, and then immediately gave them Q2 (Figure 3). That collection sequence was followed in order to prevent students from changing their Q1 answers *after* reading Q2.

The groundskeeper at the Penobscot Valley Country Club wants to know the total amount of rain (in kg H_2O) that falls on the golf course during a predicted downpour. Before the storm, four atmospheric science majors report to the golf course with timepieces and rain gauges in hand; Andy brings one rain gauge, Betsy brings four, Charlie brings 40, and Debby brings 400. You may assume all k now how to use them.

a) Which student do you expect will be able to determine a value closest to the actual total amount of rainfall? Please explain your reasoning.
b) Rank the uncertainties (as a percentage of the total amount of rainfall) associated with each student's determined value. If any uncertainties are equal, state that explicitly. Explain.

FIGURE 3. Question Q2, given in Statistical Mechanics (2006, -07) as a pretest only.

All students (N = 13) answered both parts of Q2 as expected, indicating that the relative uncertainty in the measurement of rainfall should decrease with increasing n.

We have noted the richness of statistical information available in the darts project tracking dart scores with increasing n, and yet the poor learning outcomes evident in the later *Statistical Mechanics* pretest. Consequently, we changed the darts exercise protocol in Fall 2006 (Figure 4) to add explicit individual *predictions* (based on their own scores only) for several features of the dart score distribution, emphasizing what *should change* (e.g., smoother histograms and σ_m decreasing) and what *should not change* (e.g., the distribution σ) as n (the number or reported scores) increases.

For each of the following, provide your predictions, and a brief explanation of how you made them, for:
(*i*) n = 24 (your own data); (*ii*) n = 220; (*iii*) n = 500.

a) the *average* (mean; < X >) of the n scores

b) the *standard deviation* (σ) of the n scores

c) the *standard deviation of the mean* (σ_m) of the n scores

FIGURE 4. Question Q3, given in *Physical Electronics Laboratory* (2006); these three questions are a subset of a considerably longer pretest.

TABLE 2. Results from Q3 for predicted standard deviations (σ) of dart scores for *Physical Electronics Laboratory* students (Fall 2006; N = 13)

Response	*N*
No change in σ as n increases (correct)	5
σ increases as n increases	4
σ decreases as n increases	4

The Fall 2006 results contain evidence of student confusion with statistics in their predictions from the darts project (Table 2). Of 13 students, over 60% predicted (incorrectly) that the standard deviation (σ) of the distribution *should change* with increasing n. Of those, half predicted σ would *increase* with n, while the other half predicted a *decrease* in σ. Both groups gave written explanations that they found satisfying; e.g., σ will increase with n, since "with more throws, there will be a greater spread in values…"; or σ will decrease with n "since outliers will have less and less impact…" Some of the students who predicted a decrease in σ with increasing n may be confusing σ of the distribution with σ_m, the standard deviation of the mean of the distribution (related to the relative uncertainty in Q1). However, regardless of what students indicated as a trend for σ (increasing, decreasing, or staying the same) in part (b) of Q3, they typically determined values for σ_m for part (c) by dividing their answers to part (b) by $\sqrt{n}$.

DISCUSSION AND CONCLUSIONS

Our findings so far include several that were expected, as well as some surprises on both ends of the performance spectrum. We point out some of each here, and briefly outline future plans.

Prior to instruction in statistical mechanics, essentially all of our students exhibited attributes of concepts that are consistent with appropriate models for various applications of statistics, probability and uncertainty. For example, (1) all students predicted the cumulative outcome for the number of 'heads' recorded for n coin flips to most likely be the expected value of $n/2$ [Q1], and (2) in a qualitative ranking, students agreed that the uncertainties in measurement of a continuous macroscopic quantity will decrease with n, the number of measurements made [Q2(b)].

The most serious conceptual deficiency identified here is that most of our students do *not* predict the *convergence* towards the $n/2$ (binary) macrostate with increasing probability as n increases. As seen in Q1, approximately 20% of our students start with a reasonable model for $n = 4$ coin flips (# of heads = $n/2 \pm n/2$) but remain stuck there, regardless of the magnitude of n; another 13% give smaller, but still constant, percentages of n as the uncertainty for intermediate values of n.

However, results from Q2 indicate that all our students share the widespread qualitative belief that macroscopic measurement uncertainty is indeed reduced with increasing n. Apparently, something is quite different in student understanding of the statistics describing discrete binary events [Q1] and uncertainties of measurements of continuous variables [Q2]. Even though they all have at least qualitatively appropriate models for uncertainties in a continuum context (rainfall), most fail to make a transition from discrete to continuous models as an increasing n should indicate.

Finally, the results from parts (b) and (c) of Q3 indicate that many students know the formula $\sigma_m = \sigma/\sqrt{n}$, yet lack the conceptual understanding of distinctions between σ and σ_m.

As noted earlier, the typical textbook presentation of the microscopic basis for predictable macroscopic equilibrium states via entropy and the Second Law is the dramatic convergence with increasing n to a single, well-defined, most probable macrostate of $n/2$, as seen in Figure 1. We have not yet offered curricular content that differs from such a textbook presentation in any significant way, and appropriate post-instruction student predictions for both outcomes ($n/2$) and associated uncertainties indicate that the traditional approach may be somewhat successful (Table 1). However, based on the quality and depth of the reasoning accompanying the post-instruction responses, we do not believe the conceptual difficulties with uncertainties of discrete outcomes seen with the pretests have been resolved by current instruction. Many students give correct answers supported only by terse descriptions of reasoning that are less than convincing. Since fluency with concepts of a smooth transition from the discrete to the continuous with increasing n is such a widespread and fundamental need throughout the physics curriculum, we plan to develop curricular materials to address this need, although specific instructional strategies have not yet been developed. Individual student interviews based on the written questions shown here will help us confirm common models, as well as find if students are even aware of the need for such concept development.

Informal conversations with individual students indicate one reason for incompatible rather than complementary models (discrete vs. continuous) may be that *each* binary coin flip event has zero associated uncertainty, while *each and every* measurement of a continuous variable *always* has inherent associated uncertainties. Interviews may explore further this line of reasoning.

ACKNOWLEDGMENTS

We thank an anonymous referee for useful comments. Supported in part by NSF Grant #PHY-0406764.

REFERENCES

1. J.R. Thompson, B.R. Bucy, and D.B. Mountcastle, *2005 Phys. Educ. Res. Conf.,* edited by P. Heron *et al.,* AIP Conf. Proceedings **818**, 77-80 (2006).
2. B.R. Bucy, J.R. Thompson, and D.B. Mountcastle, *2006 Phys. Educ. Res. Conf.,* edited by L. McCullough *et al.,* AIP Conf. Proceedings **883**, 157-160 (2007).
3. A.H. Carter, *Classical and Statistical Thermodynamics*, Upper Saddle River, NJ: Prentice-Hall, Inc., 2001.
4. R. Baierlein, *Thermal Physics*, Cambridge: Cambridge University Press, 1999.
5. D.S. Abbott, unpublished doctoral dissertation, North Carolina State University (2003).
6. S. Allie, et. al., *Phys.Teach*. **41**, 394-401 (2003).
7. R.L. Kung, *Am. J. Phys.* **73**, 771-777 (2005).
8. D.V. Schroeder, *An Introduction to Thermal Physics*, San Fransisco: Addison Wesley Longman, 2000.
9. A. Compagner, *Am. J. Phys.* **57**, 106-117 (1989).
10. G.L. Squires, *Practical Physics,* 2nd. ed., Cambridge: Cambridge University Press, 1976.