The Language of Conditional Probability

Jessica S. Ancker
Columbia University College of Physicians and Surgeons

*Journal of Statistics Education* Volume 14, Number 2 (2006),
ww2.amstat.org/publications/jse/v14n2/ancker.html

Copyright © 2006 by Jessica S. Ancker all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

**Key Words:** Statistics education; Statistical language.

**Abstract**

Statistical terms are accurate and powerful but can sometimes lead to misleading impressions among beginning students. Discrepancies between the popular and statistical meanings of “conditional” are discussed, and suggestions are made for the use of different vocabulary when teaching beginners in applied introductory courses.

**1. Introduction**

“I don’t know what you mean by ‘glory,’” Alice said. Humpty Dumpty smiled contemptuously. “Of course you don’t – till I tell you. I meant ‘there’s a nice knock-down argument for you!’”

“But ‘glory’ doesn’t mean ‘a nice knock-down argument,’” Alice objected.

“When I use a word,” Humpty Dumpty said in rather a scornful tone, “it means just what I choose it to mean – neither more nor less.”

Lewis Carroll, Through the Looking Glass, The Dial Press, Toronto, 1931. p. 238.

Like Humpty Dumpty, statisticians often use words to mean just what we choose them to mean. So “significant,” to a statistician, no longer means “important” but rather “unlikely under the null hypothesis.” Like Alice, students can understand the concepts only by learning to use the new language.

Such specialized statistical vocabulary can be a powerful tool for experts, but familiar terms can sometimes be better for helping beginners understand new ideas in applied statistics courses. In
particular, the tricky concept of a conditional event can be clarified if teachers avoid the unfamiliar technical phrase “A given B” and instead describe it as “A within B,” a wording that encourages an interpretation in the context of set theory. To further encourage a set-theory interpretation, teachers can describe conditional probabilities as “subset probabilities.” Introducing the ideas in easy-to-understand set-theory language can help new students focus on the important concepts and avoid several of the most common mistakes with conditional probability.

2. Common Mistakes with Conditional Probabilities

One mistake commonly made by students is to confuse P(A|B) with P(B|A), so that, for example, the conditional probability of disease given a positive diagnostic test (the positive predictive power of a test) is thought to be the same as the conditional probability of a positive diagnostic test given the disease (the sensitivity of the test). Another common mistake is failing to recognize the difference between P(A|B) and P(A): the probability of having the disease given a positive test result is thought to be the same as the probability of having the disease (the prevalence).

These mistakes stem from the same underlying problem, which is failing to recognize when the sample space – or denominator of the frequency calculation – has changed. The events A|B, B|A, and A are all thought to be the same event, and students try to describe them all with the same techniques. Effective teaching methods should thus draw attention to whatever features distinguish the events as different. The phrase “A given B” may not be sufficient to draw attention to the existence of a new category of events, perhaps because the phrase is rarely used outside of statistics. Describing conditional events as “A within B” or “A in B” can encourage students to visualize one event as a subset of another, thus emphasizing the existence of a new category. This language thus emphasizes the concept of the “reduced sample space” frequently used to teach conditional probability in probability texts (Ross 2002).

Describing conditional events as “A in B” is particularly helpful when referring to Venn diagrams or contingency tables. Both are powerful tools for visualizing event spaces and probabilities, possibly because they allow students to conceptualize them as frequencies, which are cognitively simpler to handle than probabilities and tend to lead to more accurate conclusions (Cosmides and Tooby 1996; Hoffrage, Lindsey, Hertwig and Gigerenzer 2000).

3. Problems with Thinking of Conditional Events as Sequential Events

The word “conditional” itself can be misleading for some students, because in common parlance, it usually draws attention to the fact that some event must occur before another event. For example, a student who is “conditionally accepted” to college must complete his or her high school coursework before showing up on campus. When students hear the term “conditional” used for the first time in its statistical sense, they may think of conditional events as sequential ones.

Unfortunately, conditional events in statistics sometimes become confusing if conceptualized as sequential. Consider the following probability problem from a biostatistics text (Pagano and Gauvreau 2000). United States life-table data in a certain year shows that the probability of a male birth is 0.512. If two pregnant women are chosen at random from that population, what is the conditional probability that both will give birth to boys given that we know that at least one did so?

If A is the event that at least one child is a boy, then the problem can be solved by applying the definition of conditional probability as follows:
Unfortunately, if we encourage students to focus on the sequence of events, then it appears that the sex of one child magically influences the sex of other. Even students who correctly apply the formula are likely to reject their solution as wrong because they know that for independent events such as births among randomly selected women, one event cannot affect the probability of another event. (In fact, b₁ and b₂ remain independent, although the events (b₁ ∩ b₂) and A are not.)

Set-theory language de-emphasizes the time sequence of the events by drawing attention to the fact that revealing the sex of one birth reduces the sample space. For problems like this one, the conditional probability of A given B can be expressed as the probability of A within B. The problem can be stated: “What is the probability that both children will be boys, within the subset of events in which at least one child is a boy?”

4. Subset Probabilities and P-Values

Describing conditional probabilities as probabilities in subsets is extremely useful when introducing P-values. Textbooks typically describe the P-value as “the probability of the event under the null hypothesis,” as if the null were an umbrella. However, students tend to omit the awkward “under the null” and describe the P-value as simply “the probability of an event occurring by chance.”

This rephrasing of the P-value becomes problematic if students conclude that the complement of the P-value is either the probability that the null hypothesis is false or the probability that the alternate hypothesis is true (Rossman and Short 1995). For example, some students reason, “If P = 0.01 for the difference between two means, then there’s a 99% probability that the difference is not due to chance.” When students learn about the P value as “the probability of the event in the world described by the null hypothesis,” they are less likely to be misled into thinking that its complement is the probability of the alternative hypothesis.

5. Transition to Standard Vocabulary

Teaching students to translate P(A|B) as “the probability of A within B” is useful when first introducing conditional probability in the classroom, as is describing conditional probabilities as probabilities in subsets. Using this phrasing draws their attention to the sample space and how it becomes reduced in a conditional probability. However, after students have become accustomed to this construction, then the standard vocabulary of “conditional probability of A given B” should be introduced so that students can become familiar with the language of the textbooks and the statistical literature.

6. Conclusion

“Teachers must draw out and work with the preexisting understandings that their students bring with
them,” urges one influential report on learning (Bransford, Brown and Cocking 2000). Once a teacher understands these preexisting conceptions, he or she can facilitate learning by making explicit the connections between them and the new ideas introduced in class (Knowles 1988; Lovett and Greenhouse 2000). For example, a teacher can help make Bayes’ Theorem relevant to students in medicine or other health professions by eliciting their personal experiences with false positive and false negative test results, then linking these experiences to methods for calculating conditional probability.

A teacher should consider students’ vocabulary as part of this body of preexisting understandings. Introducing conditional probability in familiar language rather than the traditional statistical terms can be a way of linking new ideas to existing knowledge, especially when the familiar language emphasizes a visual representation of the conditional events in the context of set theory. The set-theory phrasing can be used with traditional statistics texts or with a variety of elegant and useful teaching methods that have been proposed for demystifying conditional probabilities (Warner, Pendergraft and Webb 1998; Hoffrage, et al. 2000).

Students in introductory statistics classes often feel like so many Alices in a disconcerting statistics Wonderland. By using familiar language, statistics instructors can prevent this feeling by helping students seeing new concepts as relevant to the real world.

Acknowledgements

The author is supported by Robert Wood Johnson/National Library of Medicine predoctoral fellowship 3T15-LM007079-14S1 in public health informatics.

References

Bransford, J. D., Brown, A. L., and Cocking, R. R. (eds.) (2000), How People Learn: Brain, Mind, Experience, and School, Washington, D.C.: National Academy Press.

Cosmides, L. and Tooby, J. (1996), “Are Humans Good Intuitive Statisticians after All? Rethinking Some Conclusions from the Literature on Judgment under Uncertainty,” Cognition, 58, 1-73.

Hoffrage, U., Lindsey, S., Hertwig, R., and Gigerenzer, G. (2000), “Communicating Statistical Information,” Science, 290, 2261-2262.

Knowles, M. S. (1988), The Modern Practice of Adult Education: From Pedagogy to Andragogy, Englewood Cliffs, N.J.: Cambridge Adult Education.

Lovett, M. C. and Greenhouse, J. B. (2000), “Applying Cognitive Theory to Statistics Instruction,” The American Statistician, 54, 196-217.

Pagano, M. and Gauvreau, K. (2000), Principles of Biostatistics (2nd ed.), Pacific Grove (CA): Duxbury Thompson Learning.

Ross, S. (2002), A First Course in Probability (6th ed.), Englewood Cliffs, NJ: Prentice Hall.
Rossman, A. J. and Short, T. H. (1995), “Conditional Probability and Education Reform: Are They Compatible?” *Journal of Statistics Education* [Online], 3(2).
ww2.amstat.org/publications/jse/v3n2/rossman.html

Warner, B. A., Pendergraft, D., and Webb, T. (1998), “That Was Venn, This Is Now,” *Journal of Statistics Education* [Online], 6(1).
ww2.amstat.org/publications/jse/v6n1/warner.html

Jessica S. Ancker, MPH
Department of Biomedical Informatics
Columbia University College of Physicians and Surgeons
New York, NY 10032
U.S.A.
jsa2002@columbia.edu