Anirban's Angle: More Questions than Answers

Anirban DasGupta has *So. Many. Questions*. He invites your answers: send your thoughts and comments (and maybe more questions!) to bulletin@imstat.org or leave a comment on the online version at imstat.org/news.

Like a slightly impetuous, though scrupulous, young man of the teenage revolutionary kind, I am listing today an abundance of questions that more or less just occurred to me. I wish I could make at least a small fraction of *IMS Bulletin* readers spend a moment thinking about a few of them, if these questions have a nanoscopic grain of seriousness.

I do not have probable answers to most of them. Maybe you do…

1. Should PhD students in a statistics department be required to know any mathematics beyond high school algebra, two semesters of calculus, and one semester of matrix calculations? If yes, what?
2. Should statistics departments give a common degree of PhD in statistics, or separate degrees of PhD in perhaps applied statistics, data science, biostatistics, algorithms and graphics?
3. Should there be a qualifying exam for PhD students in statistics? If yes, should it be a common qualifying exam for all students, or separate exams for separate areas?
4. If there are qualifying exams, how many chances should a student get to pass it? One, two, three? Infinite?
5. Should there be some external evaluation of PhD dissertations in statistics, such as reports by anonymous experts chosen by the university?
6. Should graduate school include two semesters of ethics and integrity classes?
7. Should we teach PhD students any finite sample optimality theory for given parametric models? Or, abandon it? How soon should they learn asymptotics? Should they be required to learn Bayes theory?
8. Alternatively, should we instead start our theory classes with model-free omnibus methods, such as the bootstrap?
9. Should students see proofs? Which proofs?
10. Should PhD students in statistics be required to know more probability than probability at the Hogg and Craig level? If yes, should we mandate theory of stationary Markov chains, random walks, martingales, Brownian motion, renewals, counting processes, diffusions, large deviations, Itô theory? Which ones?
11. Should PhD students in statistics be mandated to take classes in the CS department? In the math department? In the engineering school?
12. As a policy, should hiring committees be appointed, elected, or selected by a public closed-box random selection?
13. Should promotion committees be appointed, elected, or selected by a public randomization?
14. Should admission committees be appointed, elected, or selected by public randomization?
15. Should personnel committees be appointed, elected, or selected by public randomization?
16. Should hiring decisions be made by executives, or Qualtrics surveys, or a public vote counted by a set of faculty members?
17. Generally, should tenures of department heads be limited to at most three or four years, with no renewals?
18. Should departments maintain a physical room with recent publications of its faculty on display in shelves and racks and all receptions held in that room?
19. Should annual raises pass a test of uniformity with a p-value of .10 or so?
20. How about Gallup-like tracking polls of the approval rating of department heads among faculty members and PhD students?
21. Should department heads usually teach?
22. Should citations be considered at all? If yes, how? Total number? Average? Top five?
23. Should there be a list of statistics journals in which a faculty member *must* publish as a rule to get tenured, or promoted?
24. Should faculty members who are editors be automatically given teaching reduction?
25. And, finally, should faculty members be required to make good jokes?

Since 25 is the only multiple of five below 100 that is also a perfect square, this seems to be a good place to stop.

Now this was astuteness in clean lines.

Student Puzzle Corner 33

Anirban DasGupta says this problem seems impossible at first sight, but thinks you will probably arrive at a solution quickly and it might surprise some of you that this is possible. This is the kind of thing statistics students used to learn routinely forty or fifty years ago. The problem is extremely easy to state:

(a) X is a single observation from $N(\mu, \sigma^2)$, where μ, σ are both completely unknown, i.e., $-\infty < \mu < \infty$, $\sigma > 0$. It is emphasized that the sample size in our experiment is $n = 1$. Explicitly find and plot a joint confidence region for (μ, σ) that has a coverage probability constantly equal to 0.95.

(b) Now suppose you have a sample of size $n = 2$. Derive and plot the corresponding joint confidence region for (μ, σ) that has a coverage probability constantly equal to 0.95, and find its expected area.

Deadline: September 10, 2021

Student members of IMS are invited to submit solutions to bulletin@imstat.org (with subject "Student Puzzle Corner").

The names of student members who submit correct solutions, and the answer, will be published in the issue following the deadline.

The Puzzle Editor is Anirban DasGupta. His decision is final.

Solution to Puzzle 33

A palindrome problem in probabilistic number theory: here's a reminder of the puzzle.

A positive integer is called a palindrome if it reads the same from left to right and right to left. All single digit numbers, namely, 1, 2, ⋯ , 9 are regarded as palindromes; 101 is a palindrome, or 29092, but not 111011, or 022. A zero in the first position is not allowed in the definition of a positive integer.

For $n \geq 1$, define X_n to be a randomly chosen palindrome of length exactly equal to n. For example, X_3 could be 101. Also define Y_n to be a randomly chosen palindrome less than or equal to 10^n. For example, X_3 could be 1, or 99, or 505, etc. Here are the parts of our problem.

(a) Calculate $E(X_2)$, $E(X_3)$ exactly; i.e., write the answers as rational numbers.

(b) Calculate $E(X_4)$, and then, $E(Y_2)$, $E(Y_3)$, $E(Y_4)$ exactly.

(c) Write a formula for $E(X_n)$ for a general n. Be careful about whether n is odd or even.

(d) Calculate $E(Y_8)$, $E(Y_{12})$ exactly, and recall from part (b) $E(Y_4)$.

(e) Conjecture what $E(Y_n)$ is for a general even n.

Student Puzzle Editor Anirban DasGupta explains the solution:

Well done to student members Casey Bradshaw of Columbia University, who sent a detailed, complete and correct solution; Uttaran Chatterjee of the University of Calcutta also made significant advances.

Casey Bradshaw

Uttaran Chatterjee

One first proves that the number of n-digit palindromes is $9 \times 10^{[(n-1)/2]}$ and the sum of all n-digit palindromes is $99/2 \times 10^{[3(n-1)/2]}$, where $[x]$ denotes the integer part of a non-negative number x. On division, one gets $E(X_n)$ for general n.

By summing over n the above expressions separately and then dividing, one gets $E(X_n)$. Careful summation of finite geometric series is required. For example, one will require the sums

$$\sum_{n=2}^{2k} 10^{[(n-1)/2]} = \frac{2 \times 10^k - 11}{9},$$

$$\sum_{n=2}^{2k} 10^{[3(n-1)/2]} = \frac{11 \times 1000^k - 1010}{999},$$

and the corresponding sums when the range of n ends at $2k + 1$.

It then turns out that the expected value of a random palindrome less than or equal to 10^8 is 30280280280/1111 and of a random palindrome less than or equal to 10^{12} is 302802802802802280/111111 . The pattern that emerges is clearly striking; but note that this specific pattern is for palindromes less than or equal to 10^n for an even n.