



Statistical Analysis of Exam Scores to Determine Likelihood of Cheating on a Professional Exam: A Monte Carlo Simulation

Which category are you applying in?
(Select One)

- Most Unusual

Nominee Contact Information

Alan Olinsky, Professor
Bryant University, Math/CIS
1150 Douglas Pike, Smithfield, RI
02917
401-232-6266, aolinsky@bryant.edu

Co-Author Information

John Quinn, Professor
Robert DiSario, Associate
Professor



Executive Summary

This was the result of a consultation in a court case involving accusations of cheating on a 100 question professional multiple choice exam with 4 choices for each question. As the prosecution utilized a witness who was an expert in statistical analysis, one of the authors was engaged by the defense to conduct an independent statistical analysis of the exam scores. The prosecution's witness utilized a simulation to demonstrate, in his opinion, the relative certainty of cheating. The authors performed their own analyses, including simulations to counter the testimony of the prosecution. Crystal Ball was utilized for the simulations.



Excerpts from Letter of Engagement

“Jane Doe” is a police Sergeant, who took the 1996 Sergeant’s exam. Chief “Tom Smith”, in certain Grand Jury testimony and other statements that he made, indicated that he had given “Jane Doe” the source data for the 1996 Sergeant’s exam. Chief “Tom Smith”, who was the Chief of Police at that time and has now been retired under questionable circumstances, also stated that he did so at the request of two individuals and that the Personnel Director was present at the time the request was made. All of these people have testified that this is absolutely false and that no one ever asked Chief “Smith” to give “Jane Doe” any test information. “Jane Doe” has also denied receiving the information. The Police Department appears to be making an effort to establish, through some statistical analysis, that there is a great likelihood that she cheated on the 1996 Sergeant’s exam.’



Scenario

Two individuals, now married, both received 93 out of 100 questions correct on a multiple choice examination, with 4 choices per question. They both missed the same 5 questions with the same choice of answers. They each missed 2 different questions. The exam was carefully proctored at a local community college and the chance of cheating onsite was practically nonexistent. The prosecution claimed that these 2 individuals had been given source materials. These source materials did not include questions or answers but rather the sections of the code from which the questions were selected.



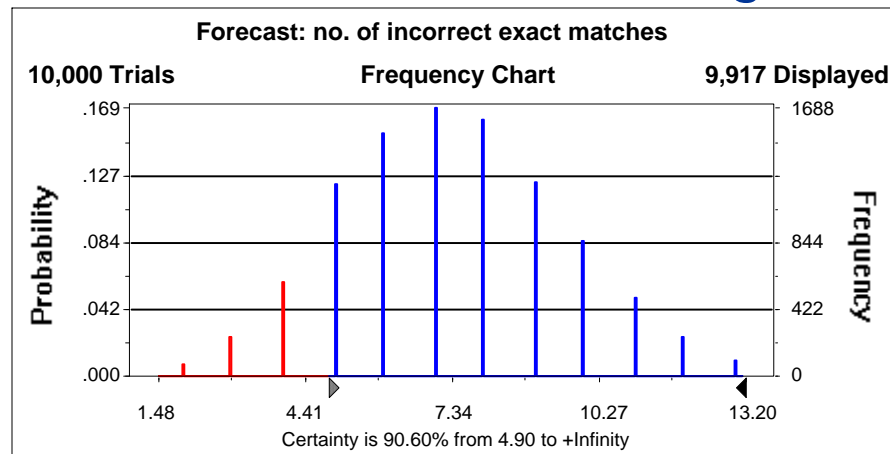
Prosecution's Expert Statistical Witness

Their expert witness testified that the probability of another test taker, under the hypothesis of no advance information, receiving 93 with 5 or more matching wrong answers with the same choices as "John Doe" was only 14 in 100,000 or .00014. He determined this probability by simulating the taking of this exam 100,000 times by the 64 exam takers using weights to match their likelihood of responding to each question with a certain choice. His conclusion was that this exam would have to be administered 7,000 times to see this kind of match. Since "Jane" had this kind of match with "John", he concluded that there was practically certain evidence of cheating.



Our Computer Simulation

Not knowing what approach would be taken by their expert witness, I, with the help of my co-authors, had independently undertaken simulations of my own. I also analyzed the class of students taking the exam and used the frequency with which they responded to the choices for each question to simulate the class retaking the exam thousands of times.



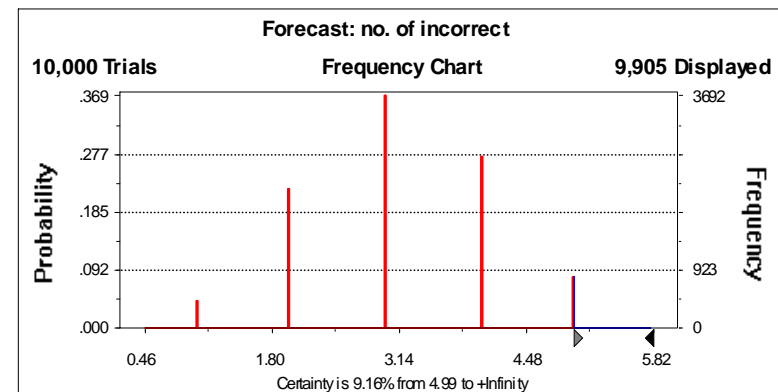
There was over a 90% chance of at least five or more matches on exact wrong answers by test takers on this exam. This is a very common event.



Computer Simulation (Continued)

We ran a simulation, in which we replicated 10,000 exams of these 64 test takers, again utilizing the actual relative frequencies with which each choice was made on each question. Although we believe that it is improper to select the individual accused of cheating for comparison with all of the other simulated students, we wanted to replicate their results. Therefore, against our better judgment, we kept the defendant's scores as a constant and ran a simulation in which we examined the number of exact wrong answers that became matches with his 7 wrong problems. In fact, after 10,000 runs, we found that approximately 10% of the class achieved at least 5 of the same wrong answers as the defendant.

Trial	Defendant's Wrong Answers	Simulated Student's Wrong Answers
1	1, 2, 3, 4, 5, 6, 7	1, 2, 3, 4, 5, 6, 7
2	1, 2, 3, 4, 5, 6, 7	1, 2, 3, 4, 5, 6, 7
...
10000	1, 2, 3, 4, 5, 6, 7	1, 2, 3, 4, 5, 6, 7





Testimony

- However, according to their expert witness's testimony, he ran a simulation that determined the probability of another test taker not only matching on at least 5 of the same wrong answers, but also achieving a 93 or higher. It should be noted that only 1 student actually had a grade higher than 93. It was in running this simulation that he determined a probability of 14/100,000 of another student matching one of the defendants. Therefore, this unlikely event, and the fact that the other individual in question achieved this, led him to his conclusion that they must have had advance information.
- To see for ourselves, we added the stipulation of matching on at least 5 of the same exact wrong answers as one of the defendants and also achieving at least a 93. The use of 100,000 is probably overkill, but we wanted to match our results with theirs. In fact, when we ran the simulation, we found no one achieving such a match.
- Although this seemed discouraging from our point of view, we noted that another student (Student 102) received a 91 on the exam and had exactly 3 incorrect matches with one of the defendants. Using the same logic as their expert witness used, we ran a simulation to determine the probability of another test taker receiving at least a 91 and having at least 3 exact matching wrong answers with the defendant. The resulting probability is also close to zero as in the simulation run by their expert. However, there was such an individual and we could use their expert's logic that these two individuals must have also had advance information and cheated. Yet there was no reason to believe that this was true and no such claim by the prosecution!



Conclusion

- We believe that this application of simulation to the determination of the likelihood of 2 students cheating by replicating the taking of an exam thousands of times with students similar to the ones at the original sitting is unique.
- As with other statistical techniques, it is a challenge to present the technique and its results to laymen sitting on a jury (in this case, 3 police officers), but it is much easier to intuitively explain a simulation and its results as opposed to a Bayesian analysis.



Conclusion (Continued)

- In conclusion, we indicated to the jury our belief that this approach used by the prosecution would be valid only in the circumstance where students are accused of physically cheating during the examination.
- Nevertheless, there are many special purpose algorithms in place to determine the likelihood of cheating when done in conjunction with the suspicion of students cheating by, e.g., sitting side by side in an exam room.
- The use of simulation has the possibility of replacing many of these techniques.
- Unfortunately, the jury found for the prosecution.
- However, on a positive note, in a subsequent appeal in Superior Court, the case was overturned and the individual was given back pay, promoted, and is back at work. 😊