Open Problems Column
Edited by William Gasarch

This Issues Column!

This issue's Open Problem Column is by Aarav Bajaj and William Gasarch and is *The Testtakers/Testmakers Dilemma.*

Request for Columns!

I invite any reader who has knowledge of an area of theoretical computer science to contact me and arrange to write a column about open problems in that area. That area can be (1) broad or narrow or anywhere inbetween, and (2) really important or really unimportant or anywhere inbetween.

The Testtakers/Testmakers Dilemma

By Aarav Bajaj and William Gasarch

1 Introduction

Alice and Bob are both taking a Biology course and are completely clueless. They have an exam and they know that they don't know anything. Alice decides to go and take the test anyway, but Bob decides to skip the test and sleep in.

The test is multiple choice with five options per question: 4 for a right answer, -2 for a wrong answer, and 0 for leaving it blank. The exam is scored that way to discourage students from guessing. Note that since Alice is clueless, for every question she guesses the expected number of points will be

$$4 \times \frac{1}{5} - 2 \times \frac{4}{5} = -\frac{3}{5}.$$

Therefore Alice should not guess at all. But emotion gets the better of her and makes some guesses. Alice gets a -8 and *Bob gets a 0 which is better!*. Professor Carol, who is fond of saying

90% of life is just showing up,

does not like seeing a student who does not show up doing better than one that does. So on the next exam Professor Carol caps the minimum score at 0. Alice notes that on the next exam she *should* guess at least one question. She has nothing to lose! Alice is terrible at Biology, but she's a mathematical genius! Hence she thinks about the following problem:

Given that an exam has:

- r is the number of points gained for a right answer,
- w is the number of points lost for a wrong answer,
- p is the probability of getting a question right,
- a large number of questions.

find G^{opt}, the number of guesses that will maximize the expected value of her score. G^{opt} is a function of p, r, and w. Let P be the expected number of points Alice will get if she guesses G^{opt} times.

Alice only cares about the case where guessing has an expected value ≤ 0, hence when:

$$rp - w(1-p) \le 0$$

Assume Alice guesses G questions. Let $0 \le i \le G$. The probability that she gets i right is $\binom{G}{i} p^i (1-p)^{G-i}$. If she gets i right and $G-i$ wrong then her score will be

$$\max\{0, ir - (G-i)w\}$$

Hence her expected value is

$$\sum_{i=0}^{G} \binom{G}{i} p^i (1-p)^{G-i} \max\{0, ir - (G-i)w\}.$$

G^{opt} is the value of G that maximizes this quantity. P is the value of this quantity with $G = G^{\text{opt}}$.

2 The SAT (the Exam, not the NP-Complete Set)

The SATs have 5 choices per question, +4 for a right answer, and -1 for a wrong answer. Hence $p = 0.2$, $r = 4$, and $w = 1$. When we first plotted this data P was an increasing function of G with no interesting modular behavior (the later examples do have such behavior). We were curious how fast the increase was, so we wanted to look at large values of G. We computed what happens for $G \equiv 0 \pmod{20}$, $20 \le G \le 600$. The data is in Table 1.

Advice for the Testmaker: With your current system of $p = 0.2$, $r = 4$, $w = 1$, you are *not* discouraging guessing! The more someone clueless guesses, the better off they are. But all is not bleak. Note that if the exam has 600 questions, and someone guesses all of them, their expected value is roughly 20. Contrast this to the max score possible of $600 \times 4 = 2400$. So, while guessing helps, it does not help much.

Advice for the Clueless Testtaker: It seems as though you should guess every question. But this might not be quite right. The more you guess the more variance there is. So your expected value is high, but if your goal is to maximize the probability that you exceed a particular score then guessing all the questions might be a bad strategy.

Lines of Research this Inspires:

1. Pin down the notion of variance to help out risk-averse clueless students.
2. What is the relation between P and G? It seems sublinear.

We expand on these lines of research in Section 8.

Table 1: $20 \le G \le 600$, $G = 0 \mod 20$

G	P
20	3.491
40	4.991
60	6.135
80	7.098
100	7.944
120	8.709
140	9.411
160	10.06
180	10.68
200	11.26
220	11.81
240	12.34
260	12.84
280	13.33
300	13.80
320	14.25
340	14.69
360	15.12
380	15.54
400	15.94
420	16.33
440	16.72
460	17.10
480	17.46
500	17.83
520	18.18
540	18.53
560	18.87
580	19.20
600	19.53

3 The University of Maryland Math Competition

The University of Maryland Mathematics Competition has 5 choices per question, +4 for a right answer, and -2 for a wrong answer. Hence $p = 0.2$, $r = 4$, and $w = 2$. The data is in Table 2.

Table 2: $p = 0.20$, $r = 4$, $w = 2$, $1 \leq G \leq 20$

G	P
1	0.8000
2	0.9600
3	0.6720
4	0.8960
5	0.9152
6	0.7050
7	0.8225
8	0.8142
9	0.6517
10	0.7241
11	0.7079
12	0.5790
13	0.6273
14	0.6094
15	0.5055
16	0.5392
17	0.5219
18	0.4373
19	0.4616
20	0.4457

Advice for the Testmaker: With your current system of $p = 0.2$, $r = 4$, and $w = 2$, you have discouraged guessing. I am sure you are happy that a clueless guessing student, if they are good enough at math to figure out that $G^{\text{opt}} = 2$, has an expected score of only 0.96. The UMCP Math competition has 25 questions, so the max score is 100. Hence 0.96 is ... not much.

Advice for the Clueless Student: Guess $G^{\text{opt}} = 2$ questions and study more.

Lines of Research this Inspires: We originally thought that $P(G)$ would increase and then decrease. *This did not happen!* But the reader can check our data and find that if G is restricted to 0 mod 3 (or 1 mod 3, or 2 mod 3) then $P(G)$ *does* increase then decrease. In addition, $P(3x + 1) > P(3x)$. We will see a mod pattern again later in Section 4. We will make a conjecture about mod patterns in Section 7.

4 Dr. Gasarch's Formal Language Theory Final

On his final in Formal Language Theory, Dr. Gasarch often has a question where he gives several languages and for each one asks if it is (1) Finite, (2) Regular but not Finite, (3) Context Free but not Regular, (4) P but not Context Free, (5) NP but not P (the students may assume P≠NP for this problem), (6) Decidable but not in NP, (7) Computably enumerable but not Decidable, (8) Not computably enumerable.

He gives +5 for a right answer and -2 for a wrong answer. Hence $p = 0.125$, $r = 5$, and $w = 2$. The data is in Table 3.

Advice for the Testmaker: With your current system of $p = 0.125$, $r = 5$, and $w = 2$, you have discouraged guessing. I am sure you are happy that a clueless guessing student, if they are good enough at math to figure out that $G^{\text{opt}} = 2$, has an expected score of only 0.8125. Dr. Gasarch typically asked 10 questions of this type, so the max score was 50. Hence 0.8125 is... not much. In addition, there other questions on the exam, so if a student got any of them right, they should not guess.

Advice for the Clueless Student: You should have dropped the course after the midterm.

Lines of Research this Inspires:

Is there a pattern to the data? At first glance no. But look at the data with G mod 7 and one realizes the following

1. The testtaker is better off guessing $7x + 1$ questions rather than $7x$.
2. The testtaker is better off guessing $7x + 5$ questions rather than $7x + 4$.
3. For $x \geq 2$, of $7x, 7x + 1, \ldots, 7x + 6$, the best number to guess is $7x + 1$.

We will make a conjecture about this modular pattern in Section 7.

Table 3: $p = 0.125$, $r = 5$, $w = 2$, $1 \le G \le 28$

G	P
1	0.625
2	0.8125
3	0.6445
4	0.5244
5	0.6032
6	0.5590
7	0.3716
8	0.4247
9	0.4226
10	0.3468
11	0.2929
12	0.3044
13	0.2773
14	0.1997
15	0.2140
16	0.2078
17	0.1732
18	0.1482
19	0.1504
20	0.1367
21	0.1016
22	0.1064
23	0.1025
24	0.08619
25	0.07424
26	0.07452
27	0.06772
28	0.05124

5 True-False Tests

In Table 4 we look at True-False tests, so $p = 0.5$, with $r = w = 1$. Notice that there is a pattern. We leave it as an open question to determine the pattern and prove it. The reader is warned that the numbers in the table are rounded, and hence not exact. Note that there is a mod 2 behavior here, however,

The advice and lines of research for this are similar to that of the SAT exam. We especially note that $p = 0.5$ and $r = w = 1$ does not discourage guessing, but it does limit how profitable guessing is.

Table 4: $p = 0.50$, $r = 1$, $w = 1$, $1 \le G \le 30$

G	P
1	0.5
2	0.50
3	0.750
4	0.7500
5	0.9375
6	0.9375
7	1.094
8	1.094
9	1.230
10	1.230
11	1.354
12	1.354
13	1.466
14	1.466

6 Cases Where G^{opt} is Large But Not Infinite

In the examples above either G^{opt} was infinity or 2. This is because we either had expected value of a question is 0 (so guessing is a good idea) or a harsh penalty (so guessing is a very bad idea). If the penalty is not that harsh then (not surprisingly) there are cases where G^{opt} is large but not infinity. The following table gives some of those cases by holding $p = 0.20$, and $w = 1$, but varying r from 3.0 to 3.8.

Table 5: $p = 0.2$, $3.0 \le r \le 3.8$, $w = 1$

r	G^{opt}	P
3.0	22	1.404
3.1	31	1.622
3.2	40	1.894
3.3	58	2.247
3.4	77	2.719
3.5	124	3.386
3.6	200	4.389
3.7	364	6.067
3.8	804	9.423

7 The Modular Conjecture

In Sections 3 and 4 we saw some interesting modular behavior. We make a conjecture that we are not sure we believe. We are more asking for *when* it is true.

Conjecture 1. Let p, r, w be such that $rp - w(1-p) < 0$ (so guessing would normally be a bad idea). Then there exists M such that the following occur:

1. For $0 \le i \le M-1$, for large G, if you look at $G \equiv i \pmod{M}$, then $P(G)$ goes up and then down (it may just go down).
2. For large x, guessing $Mx+1$ is better than guessing $Mx, Mx+2, \ldots, Mx+M-1$.

8 Lines of Research When Guessing Normally Has Expected Value 0

We reiterate and expand on the lines of research that the SAT exam and True-False exam inspired. Assume that

$$rp - w(1-p) = 0 \text{ (so guessing would normally have expected values 0).}$$

Let $P(G)$ be the expected number of points if there are G guesses.

1. We suspect that $P(G)$ is strictly increasing, and that this should not be too hard to prove. (Why didn't we? Because the authors are as good at math as Alice and Bob are at Biology.)
2. Of more interest: It seems as though $P(G)$ is sublinear. We conjecture that this is true and wonder how to best describe it.
3. What if Alice wants to maximize her probability of getting a passing grade? Then she might not be well off guessing *every* question. Table 6 shows what happens when $p = 0.2$, $r = 4$, $w = 1$, $x = 5$ (this is the passing grade), and Q is the probability that Alice gets a score > 5. Note that there is a tendency for more questions leading to a higher probability, but there seems to be modular effect. Let $Q(G)$ be the probability of getting > 5 if Alice guesses G questions. It looks like

   $$Q(5k+1) < Q(5k+2) < Q(5k+3) < Q(5k+4) < Q(5k) < Q(5k+5).$$

   Hence Alice should guess the largest G that is a multiple of 5.

   As G goes to infinity does the probability of passing goto 1 or goto some limiting value < 1? The data in Table 7 indicates a limiting value < 1.

   Based on this scant evidence we make the following bold conjecture.

   Conjecture 2. Let p, r, w be such that $rp - w(1-p) = 0$. Let $x \in \mathbb{R}^+$. Let $Q(G)$ be the probability that, if Alice guesses G questions, she gets $> x$ on the exam.

   (a) Then there exists M such that, for large k, for $1 \le i \le M-2$:

   $$Q(Mk+i) < Q(Mk+i+1) < Q(Mk) < Q(Mk+M).$$

   (b) There is a value $\alpha < 1$ such that the limit of the probability of passing is α.

Table 6: $p = 0.2$, $r = 4$, $w = 1$, $x = 5$

G	Q
1	0
2	0.04000
3	0.1040
4	0.1808
5	0.2627
6	0.09888
7	0.1480
8	0.2031
9	0.2618
10	0.3222
11	0.1611
12	0.2054
13	0.2527
14	0.3018
15	0.3518
16	0.2018
17	0.2418
18	0.2836
19	0.3267
20	0.3704
21	0.2307
22	0.2674
23	0.3053
24	0.3441
25	0.3833
26	0.2526
27	0.2866
28	0.3216
29	0.3571
30	0.3930

Table 7: $p = 0.2$, $r = 4$, $w = 1$, $x = 5$

G	Q
100	0.4405
200	0.4578
300	0.4655
400	0.4701
500	0.4733
600	0.4756
700	0.4774
800	0.4789
900	0.4801
1000	0.4811

9 Acknowledgments

We would like to thank Rob Brady, Clyde Kruskal, and Prahalad Rajkumar for proofreading and helpful discussions.