David Roth

Professor Zambrano

Econ 409

4 June 2025

Grade Optimization

Ⅰ Summary

In this paper, I develop a model to determine the optimal amount of time a student needs to study, taking into account random variables that affect exam scores. To help set the model (figuring out number scaling, the effects of random variables, etc.), I received assistance from second-year peers, who were mostly students in STEM fields. In the paper, I will first build the general model—model without numbers—then add values. Finally, I will delve into the optimal amount of hours to study for each class given key assumptions and subjective student beliefs.

Ⅱ Introduction to the Problem

Foremost, college students attempt to achieve a certain GPA. Doing so requires a certain level of performance on assessments. Students, especially in STEM or more technical fields, have many assessments that dictate their GPA.

At California Polytechnic State University, which is on the quarter system, students typically take four classes (16 units), each usually consisting of some exams. To achieve the academic excellence needed to satisfy one's own GPA requirement, students need to study for their exams. However, most students do not want studying to become a large ordeal of their student lives and would rather find a balanceable level of time studying and time spent on leisure activities.

Thus, assuming that all exams matter equally to the student and the student is taking 16 units, the problem boils down to: how many hours do I need to study for each exam during the week to get certain grades on the exams most of the time, while trying to maximize one's leisure. This problem will be analyzed by examining key percentiles, which students dictated matters to them. Furthermore, we will consider leisure as anything other than sleep, doing homework, going to class, and studying.

Ⅲ **Structure of Model**

Time Constraint

Foremost, I assumed that students would all have the timeframe of a week to spend studying, with all exams occurring on the same day. Thus, the student`s time constraint looks as follows:

$$168 = t_s + s + t_c + l + h$$

where t_s is time studying, s is time spent sleeping, t_c is time going to class, l is leisure, and h is homework.[1]

Also, I assume that a student is going to every class and is sleeping for eight hours each night. Further, I assume the student has around 15 hours of homework a week. Thus, the time constraint becomes one of two unknowns[2]:

$$81 = t_s + l$$

Grade Equation

To make a determinant model of a student's grade on an exam, I devised a formula that consists of the number of hours the study studies for the class t_{s_i}, where $i \in [1, 4]$,[3] the efficiency of a student e, and a scaling coefficient designated as the student's ability a. The efficiency of a student e will be assumed to be fixed such that e equals a constant $\bar{e} \in (0, 1)$, since studying should show diminishing returns. Thus, the determinant model of a student's grade used in this paper is[4]:

$$G_{student_i}(t_{s_i}) = a(t_{s_i})^{\bar{e}}$$

---

[1] The 168 stands for the number of hours in a week.
[2] Since the student is assumed to be taking 16 units, which corresponds to around 16 hours a week of class, 15 hours studying, and 56 hours sleeping, time in the week is 168 - 16 - 15 - 56 = 81 hours.
[3] i = 1 refers to class 1, i = 2 refers to class 2 and so on.
[4] Inspiration in devising Cobb-Douglas esque model comes from models developed in:
Bratti, Milena, and Andrea Staffolani. "Student Time Allocation and Educational Production Functions." Journal of Economic Studies, vol. 29, no. 1, 2002, pp. 3–18.
Douglas, David C., and James M. Sulock. "Estimating Educational Production Functions with Sample Selection." Economics of Education Review, vol. 14, no. 1, 1995, pp. 1–15.
Salemi, Michael K., and George Tauchen. "Estimation of Nonlinear Learning Models." Journal of Econometrics, vol. 18, no. 1, 1982, pp. 59–74.

Random Variables

To figure out random variables that plague students on exams, I asked peers and found that most people I asked said there was the number of “stupid” mistakes and the difficulty of an exam.

For “stupid” mistakes, these mistakes account for the following types of mistakes and more: reading questions wrong, writing down information wrong, arithmetic errors, filling in the wrong bubble in a multiple choice test, etc. However, “stupid” mistakes do not account for making mistakes on an exam or getting a question wrong due to a lack of knowledge or bad study habits.

Then, there is exam difficulty. Exam difficulty is two parts. First, it is the notion that before one takes the exam, they are unsure of the difficulty of the questions on the exam. Also, exam difficulty accounts for the “difficulty” of the grader, which we are assuming in this paper is unknown. Thus, students identified that there were easy exams—exams that had questions easier than anticipated and/or had far lenient grading than expected—normal exams—average difficulty exams—and hard exams—exams with very tough questions and/or very tough rigorous grading.

To model mistakes and difficulty, I presented several distributions to students, seeing which distributions students believed most accurately captured their subjective beliefs about the number of mistakes they make and the difficulty of exams run into during their time at college. For mistakes, most students believed the lognormal distribution fit best, and for difficulty, students were split on a normal distribution and a triangular distribution. To set the distribution for difficulty, I took the two distributions—normal and triangular—added the sum of the proportions in favor for each distribution and divided by the proportion in favor of the triangular and normal distribution, getting 0.2727 and 0.7273.

For the rest of the paper, I will refer to exam difficulty as $\underline{D}$ and “stupid” mistake as $\underline{M}$, where the bar represents the fact that these are random variables.

Accounting for Class Difficulty and Student Effort

Students, when asked, said they typically try to balance out the course load taken during each academic term to help achieve a certain level of academic rigor. Thus, to account for class difficulty, I assumed students would take four classes—two being equally hard classes and two being regular difficulty classes.

Also, students believed that the more effort put into the exams, the less randomness affects them. I.e., the more study time I put in, the less the difficulty of the exam/grader should affect me and the number of “stupid” mistakes should overall decrease. To account for this, I added an equation assuming diminishing marginal returns:

$$IF(t_{s_i} > t_{threshold}, \sqrt{t_{s_i} - t_{threshold}}, 0)$$

where $t_{threshold}$ is the time needed that once one studies an amount more than $t_{threshold}$, there is a positive concave decrease (diminishing marginal returns) in the effects of randomness on the grade for class i. The paper will refer to studying such that $t_{s_i} > t_{threshold}$ as studying in excess.[5]

Model without Values

Therefore, the equations used in the model are:

$$81 = t_s + l$$

$$G_{student_i}(t_{s_i}) = a(t_{s_i})^{\bar{e}} + \underline{D} + \underline{M} + \begin{cases} \sqrt{t_{s_i} - t_{threshold}} & \text{if } t_{s_i} > t_{threshold} \\ 0 & \text{if } t_{threshold} \geq t_{s_i} \end{cases}$$

$$t_s = t_{s_1} + t_{s_2} + t_{s_3} + t_{s_4}$$

[5] The square root function was chosen since it met the requirements of a positive concave function.

Ⅳ Adding Values and Distributions to the Model[6]

First, I needed to find values for the distributions. To do so, I asked individuals for the first, second, and third quartiles for $\underline{M}$ and the min, mode, mean, max, and standard deviation for $\underline{D}$. To see if there was any correlation between the two, I simulated $\underline{D}$ for around 1000 simulations and took the quartiles. Then, I asked each student for their conditional median using $MIDRAND(q, p)$ to find how students` values to $\underline{M}$ would change based on the quartiles for $\underline{D}$. Once I asked each student, using $CORREL$, I found each correlation, averaged out the correlations, and then used $CORAND$.

For $\underline{D}_{regular}$, I got the min, mode, and max to be -10, -4, 8, respectively, a mean of -4.15 percentage points, and a standard deviation of 4.5 percentage points. For $\underline{D}_{hard}$, I assumed around 5 points difference and subtracted 5 from the min, mode, mean, and max. Thus,

$$\underline{D}_{regular} = IF(RAND() < .2727, TRIANINV(\underline{A_1}, -10, -4, 8), NORM.INV(\underline{A_1}, -4.15, 4.5)).$$

Note: $\underline{A_1}$ is from $CORAND()$ and is the same value in both $TRIANINV$ and $NORM.INV$ and is correlated with what will be denoted as $\underline{A_2}$.

For $\underline{M}_{regular}$, I averaged out the quantiles after adjusting for the correlation and got 2.5 percentage points for quartile one, 4.25 percentage points for the median, and 7.5 percentage points lost for quartile three. (For $\underline{M}_{hard}$, I added 2 points from each quartile.) So,

$$\underline{M}_{regular} = -1 * GENLINV(\underline{A_2}, 2.5, 4.25, 7.5).$$ [7]

Finally, there is the scaling of the Grades function. To do this, I needed the value of t_{s_i}, $\bar{e}$, a, and an appropriate grade. To determine a basis for t_{s_i} I asked students for the number of hours studied for a regular difficulty class, took the average, and then rounded to the nearest

---

[6] The values for the essay were dependent on student subjective beliefs. I collected data from several peers and then took the mean of the responses, consistently consulting with the peers to make sure the aggregate numbers were close to how each individual interpreted the values.

[7] The negative one accounts for the values being percentage points lost.

half hour. Thus, the initial basis for t_{s_i} was 5.5 hours, which I also set for $t_{threshold}$. Then, I assumed arbitrarily that $\bar{e}$ would be 0.8. I gathered that most students I asked had a GPA of 3.7 or higher, so assuming that the entire grade of a class is not exams, I set Grades function to be 87% and solved for a. Therefore, $a = 24.2$. To stop exam grades from being less than 0 or more than 100, I used $MAX()$ and $MIN()$ respectively.[8]

Ⅴ Implementation

Taking all the finalized equations, I ran 5000 Monte Carlo simulations in Excel for the points lost for each exam. So, one column was points lost in exam 1, then points lost in exam 2, and so on. Several rows over was the formula for calculating grades all based on the same cells for time. I.e. one cell had time for exam 1, one for exam 2, and so on, where each one of those cells controlled the number of hours studied. Then, in the following cells, there are computed statistics and the time constraint.

Ⅵ Results

Returning to the goal of the paper, this section will cover the steps taken to find the optimal solution for sampled students. I asked each student what percent of the time they want to achieve a certain grade on an exam. On average, they wanted to obtain some form of an A (A or A-) on the exam around 90 percent of the time. Thus, using the grades from the monte carlo simulation and $PERCENTILE.INC$, I obtained each exam's tenth percentile. Then, I had the following optimization problem:

$$\max_{\ell} 81 = t_s + l$$

$$G_{student_i}(t_{s_i}) = a(t_{s_i})^{\bar{e}} + \underline{D} + \underline{M} + \begin{cases} \sqrt{t_{s_i} - t_{threshold}} & \text{if } t_{s_i} > t_{threshold} \\ 0 & \text{if } t_{threshold} \geq t_{s_i} \end{cases}$$

$$PERCENTILE.INC(Simulations_i, 0.10) \geq 90$$

[8] I took the max of grades and 0, stopping negative grade values, and took the min of grades and 100, stopping grades from being above 100.

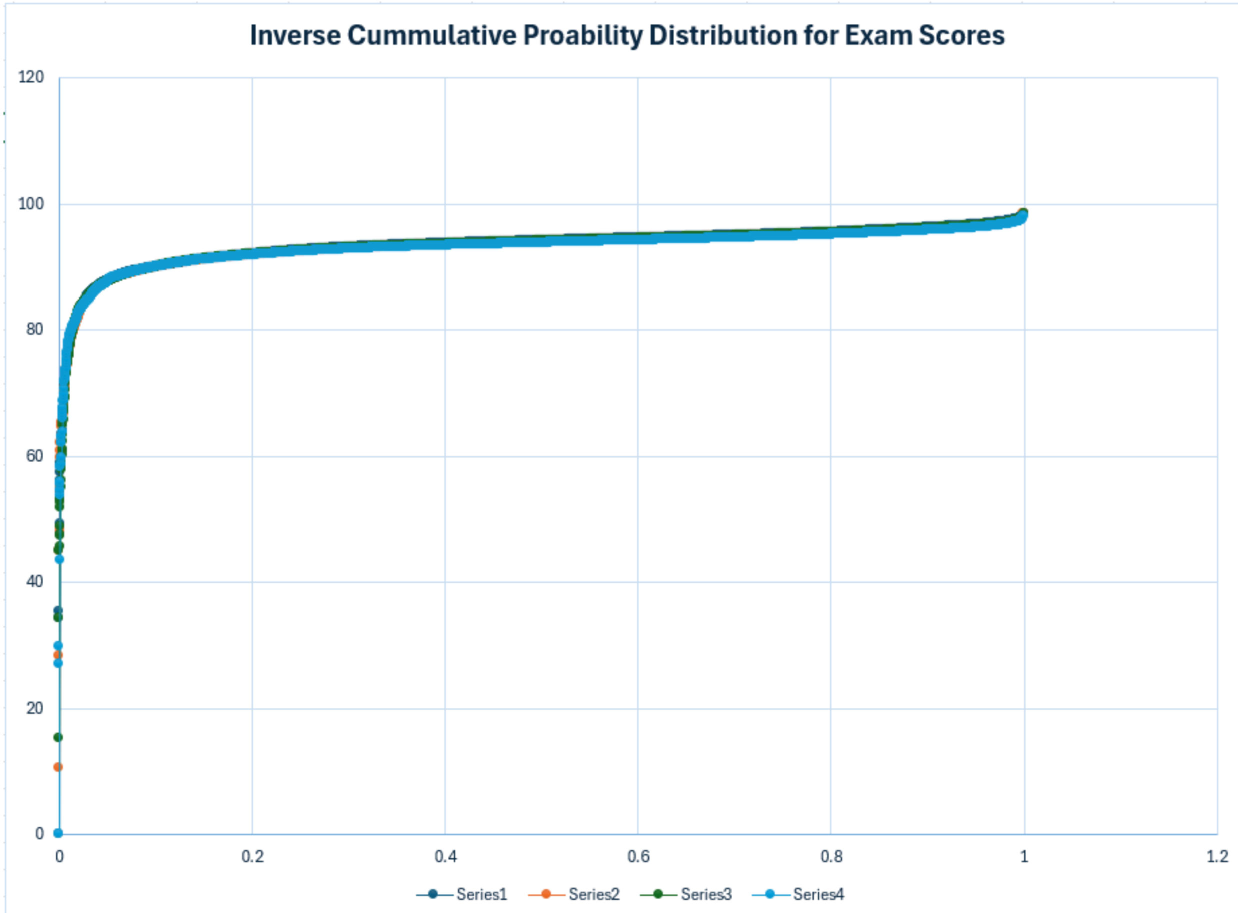
Using Solver, I maximized leisure, subject to getting the percentiles all to be greater than or equal to 90 for each exam, via changing t_{s_1}, t_{s_2}, t_{s_3}, t_{s_4}. I obtained a solution, where $t_{s_1} = 6.57\, hours$, $t_{s_2} = 6.56\, hours$, $t_{s_3} = 6.06\, hours$, and $t_{s_4} = 6.04\, hours$. Thus, $t_s = 25.23\, hours$, so $l = 55.77\, hours$. Therefore, to obtain the exam scores wanted 90 percent of the time, the students would need to study in excess of around an hour for hard classes and half an hour for regular classes.

To examine the risk for the 10% of the time when the student falls short of an A-, I calculated the expected shortfall associated with each exam. For each exam, the expected shortfall was around 84%. So, the expected loss is 6 percentage points lower than the tenth percentile.

Comparing these results to the 10th percentile of studying 5.5 hours for each class, by studying an hour more, students increase both the 10th percentile and expected shortfall by around 14 percentage points for hard classes. For regular classes, studying half an hour more increases the 10th percentile and expected shortfall by around 8 percentage points.

However, there are some worst-case scenarios worth mentioning. Notably, for hard classes, there is a 0.56-0.58% chance of getting below 70% on the exam, and for regular classes, there is a 0.56-0.66% chance of getting below 70%. Also, there is a very slim chance, less than 0.02%, of getting a zero on the exam.

Thus, students with these characteristics and the assumptions stated throughout the paper, according to the simulations, should study around 25 hours a week and then have around 56 hours leftover for leisure.



Ⅶ Usage of Model and Improvements

Since all the numbers in the simulation are based upon a student's subjective probability distributions, a student can take this model, modify key assumptions made such as the number of hard classes or regular classes being taken and calculate the amount of studying needed to achieve their academic desires.

For improvements to the model, the efficiency, e, may decrease as studying gets done, resulting in further depreciating returns in grades. Also, some numbers, such as $t_{threshold}$ were chosen based on convenience, and the formula for excess studying for instance was simply chosen because it met a positive concave function.

Works Consulted

Bratti, Milena, and Andrea Staffolani. “Student Time Allocation and Educational Production Functions.” Journal of Economic Studies, vol. 29, no. 1, 2002, pp. 3–18.

Douglas, David C., and James M. Sulock. “Estimating Educational Production Functions with Sample Selection.” Economics of Education Review, vol. 14, no. 1, 1995, pp. 1–15.

Salemi, Michael K., and George Tauchen. “Estimation of Nonlinear Learning Models.” Journal of Econometrics, vol. 18, no. 1, 1982, pp. 59–74.