Could my father graduate from PoliTo?
Computer-aided simulations lab
Academic Year 2023/24
Homework G5

Marcelo Bastos Lopes Ferreira - S308964

I. Simulator description

The simulator described and tested during this report aims to answer mainly the following questions:

- My father, also an engineer, always says he would easily graduate from the course I am in, but would he?
- Can we answer the previous question with a simulation, while also trying to predict his average grades, graduation times, and dropping/getting expelled probability, in case the course existed when he graduated engineering?

Jokes aside, the objective of the simulator is to answer the following question, for the Data Science and Engineering course from Politecnico di Torino:

- How much does the year influence the students' average grades, graduation times, and dropping/getting expelled probability?

A. Input Parameters and Assumptions

1) Simple parameters: I will start off with the simplest input parameters and assumptions, which we can get from the real Data Science and Engineering course statistics, or reasonable assumptions for the values based on personal experience: $Output\ Metrics$:

- number of subjects of the course (fixed to 13)
- number of sessions per year (fixed to 6)
- seed (fixed to 42)
- maximum number of exams per session (fixed to 4)
- average number of exams taken per session (fixed to 2)
- confidence level for the confidence intervals (fixed to 0.99)
- minimum accuracy acceptable (fixed to 0.99) on the confidence intervals of the output metrics to stop running the simulation, with the confidence level specified above
- the minimum grade on an exam that a student will accept (fixed to 18)
- maximum number of years a student can get to graduate before being expelled/dropping (fixed to 4)
- maximum number of subjects a student can pass in a year (fixed to 8)
- year (variable)

2) From year to average IQ: The IQ distribution of the general population is normalized to have an average of 100 and a standard deviation of 15 (normal distribution), but this normalization has to, every year or so, be re-normalized because the average IQ of the population has been shown to increase over the years, and this effect has been denominated as Flynn effect (the variation of the standard deviation is still up to debate so we will assume it is constant).

In the meta-analysis [1], they make available how much the average IQ has shifted (approximately) every year, all the way back from 1924 to 2013. So, for our purposes, I fit a third-order polynomial on the data to get a smooth function, to get the average IQ of a population according to the year, and then normalized the function so that it would give an average IQ of 100 to 2023.

For example, in 1990 (when my father did his engineering bachelor), the average IQ of the population would be 94.27 according to today's scale. The visualization of the smooth approximation and the data from the paper can be seen in Fig. 1.

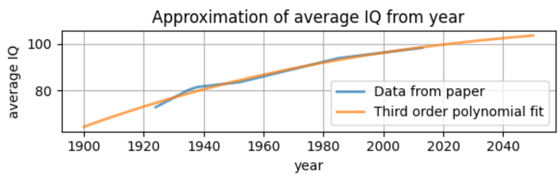

Fig. 1: Approximation function from year to average IQ of the general population

3) From IQ to average grade: It has been long known that IQ is very related to grades. In [2], they show a second-order polynomial regression capable of predicting the IQ of a student given the SAT score (USA national level exam) with a correlation of 0.857 and standard error of 5.94. The predictive equation is:

$$IQ = (0.126 \times SAT) + (-4.71e{-}5 \times SAT^2)) + 40.063$$

But this equation cannot be simply inverted to give a relationship from IQ to SAT score because its maximum value is at 1337, when the maximum score for the SAT is 1600, meaning the function is not invertible because it is not injective in its domain.

The solution for this problem, used in this report, is to fit a linear equation $(ax+b)$ on the local inverse between 400 and 1337 of the equation given in the report.

After getting this first-order linear approximation, the equation was also normalized so that the range of possible values from the SAT are mapped to possible grades in PoliTo, that is, [0, 30]. This also means that any IQs that would lead, according to the approximation function, to average grades greater than 30 or smaller than 0 are clipped to be inside the possible range. This would just mean that students with an IQ over 145 would have an average of 30 at PoliTo, which is fine because according to general knowledge, people with IQ over 145 are considered geniuses.

Mathematically, our function that maps an IQ into an average PoliTo grade is:

$$Grade_{avg} = \min(\max(0.53 \times IQ - 47, 0), 30)$$

B. Stochastic elements

The number of exams taken in a session is $\sim Binomial(n, p_{bin})$, where:

- $n = maximum\ number\ of\ exams\ per\ session$
- $p_{bin} = \frac{average\ number\ of\ exams\ taken\ per\ session}{maximum\ number\ of\ exams\ per\ session}$

The engineering student's IQ is $\sim Normal(IQ_{avg} + 26,\ 15)$, where IQ_{avg} is the one calculated as explained in the *Input Parameters and Assumptions* section from the year of the simulation. The "+26" is there because, according to the data found here, engineering graduates have an average IQ of 126, so we will assume that we can translate the IQ distribution by just adding 26 points to the mean.

Making the assumption that exams at PoliTo have 31 exercises, each one with a value of 1 point, we can model each exam's grade as $\sim Binomial(n, p)$, where:

- $n = 31, \quad p = \frac{Grade_{avg}}{n}$

And, in this way, even though we defined that students could have a maximum grade average of 30, they can still get a few lodes, as $n = 31$.

C. Output metrics (for the simulation of one year)

When the simulator finishes running, that is, when the following averages have acceptable accuracy, as described in the *Inputparameter* subsection, it will return a data object with all necessary data values to compute the following metrics, which are printed/plotted at the end of the simulation run, as requested in the padlet:

- average final grade (where final grade is the average of a student's grade, rounded to the closest integer between 18 and 30)
- average graduation time
- average number of sessions taken to graduate
- histogram of graduation times
- histogram of final grades
- plot of graduation time (x-axis) by the final grade (y-axis)

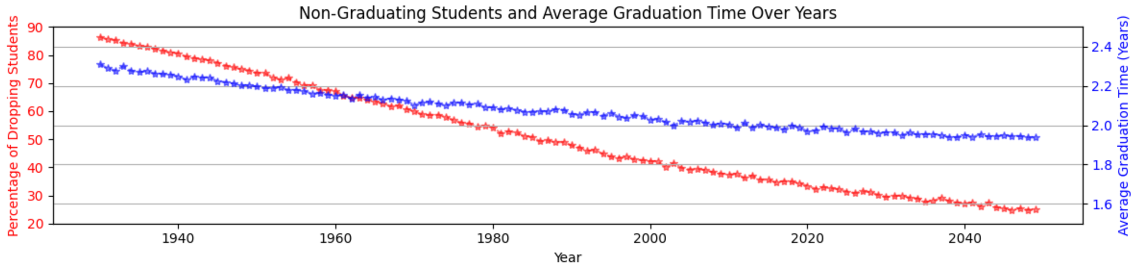

Fig. 2: Time taken to graduate and probability of not graduating for variable year

II. Experiments

A. Varying year

In this first experiment, shown in Fig .2, it is possible to see that when increasing the year, as expected, the time to graduate decreases as well as the probability of not graduating. Of course, if the maximum number of years to graduate was higher, the average graduation time would increase faster when ”going back in time”, and this hypothesis can be supported by the fact that the number of non-graduating students decreases approximately 5 percentual points per decade.

B. Years 1990 and 2023

	2023	1990
Dropping Probability	0.3250	0.4781
Average Grade	24.2225	23.4397
Average Graduation Time	1.9846	2.0567

TABLE I: Comparison of students for 2023 and 1990

Unfortunately, as can be seen in Table .1, my father (who graduated his bachelor's in engineering in 1990) is right and he would indeed (with a probability of 0.4781) be able to graduate from the Data Science and Engineering course at Politecnico di Torino.

And, by comparing directly the histograms of final (average) grades and graduation times shown in Fig. 3, we can see that grades are a bit shifted more toward higher grades, and the number of years to graduate, as expected, is shifted a bit towards the lower number of years.

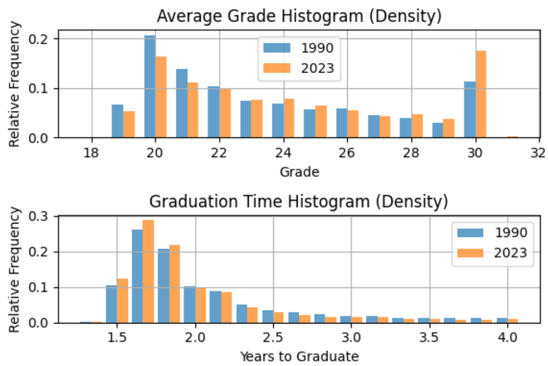

Fig. 3: Comparing graduation times and average grades between students from 1990 and from 2023 taking the same course

III. Final comments

The plots shown and discussed in the *Experiments* section seem to be a reasonably accurate description of what happens at Politecnico di Torino in real life (from personal experience), and that, consequently, shows that the simulator probably works well.

Even the behavior that more students have 30 than most of the other grades have been correctly transferred into the simulator.

When running the simulator, note that it does not automatically return all the figures shown during this report, but rather, prints out all the input and output parameters for the year of choice, and also plots the ones asked for in the padlet. Also, note that no complex data structures were needed for the simulator, and the output metrics were stores as summations, that is, singular values.

REFERENCES

[1] J. Pietschnig and M. Voracek, "One century of global iq gains: A formal meta-analysis of the flynn effect (1909–2013)," *Perspectives on Psychological Science*, vol. 10, no. 3, pp. 282–306, 2015.

[2] M. C. Frey and D. K. Detterman, "Scholastic assessment or g? the relationship between the scholastic assessment test and general cognitive ability," *Psychological science*, vol. 15, no. 6, pp. 373–378, 2004.