Underperformance on the Mathematics SAT exam correlates with being underrepresented in the STEM workforce

Ivan Kramer
University of Maryland Baltimore Country (United States)

Correspondence: kramer@umbc.edu

Published by Psychreg Ltd
ISSN: 2515-138X

A new bell curve is introduced that is a more accurate representation of a dataset than the Gaussian function. Fitting our new, naturally saturated, representation of the percentile function of a dataset, our new bell curve is simply the slope of this curve as a function of any point in the finite range of the dataset. Excellent results are shown to be obtained when this representation is applied to the 2021 Math Scholastic Aptitude Test (SAT) scores for African American, Hispanic, White, and Asian populations. Mathematics is a gateway subject for succeeding in the science, technology, engineering, and mathematics (STEM) fields. In fact, African Americans and Hispanic college graduates and under-represented among degree recipients in STEM fields while Whites and Asians are over-represented. Unsurprisingly, therefore, African American and Hispanic workers are under-represented in the STEM workforce compared with their share of all workers, while Whites and Asians are over-represented. These facts suggest that Math SAT scores correlate with subsequent success in obtaining and holding STEM jobs.

Keywords: bell curve; college students; dataset; Gaussian function; mathematics

The distribution of values v in a dataset is best represented by a percentile distribution function, to be denoted by PD(v) here, which varies between 0 and 1 (100%). The derivative of the PD(v) with respect to v is a bell curve since it represents the probability per unit value of obtaining the value v in the distribution. The PD(v) function that best fits the dataset will yield a bell curve that is superior to all others. The range of values for v for a typical dataset is proscribed, having theoretical minimum and maximum values. Exams, for example, are typically graded with the fractional score of 0 to 1 (0% to 100%), so v ranges from 0 to 1. The main problem with the gaussian bell curve is that it vanishes only at plus or minus infinity, not at 0 and 1 (100%). The PD(v) function to be introduced here naturally saturates at the finite minimum and maximum values for v and produces a bell curve that is a superior fit to the data than the Gaussian function.

The best way to introduce our naturally saturated percentile distribution function PD(v) is to fit it to the data of the concrete example with. Consider the distribution in the duration of human gestation in a cohort of pregnant woman whose deliveries are spontaneous. Such data was published paper by Jukic et al. (2013) and is reproduced in Figure 1a. In this paper, the percentile distribution function that fits this dataset will be represented by

$$(1). \qquad PD(d) = [1 - exp(-\frac{d}{C(D-d)})]^n,$$

where d is the number of days after ovulation, D is the maximum duration of any pregnancy, and n is a positive number that helps determine the sharpness of the associated bell curve. Notice that this function vanishes at the minimum value of *d* = 0 and the maximum values d = D days, as it must on physical grounds, and is therefore naturally saturated at the endpoints of its physical range. Notice also that the gaussian function does not satisfy these boundary requirements.

Using a least-squares programme to fit the function in Equation 1 to the dataset in Figure 1a produces the excellent result also shown in Figure (1a). The derivative of this fitted PD(d) function with respect to d gives our model bell curve shown in Figure 1b, a function that peaks at its most probable value of d_{mp} = 270 days. This modelling result is consistent with Jukic's calculations of the average ovulation-based gestation of 268 days (standard deviation (SD) of 9 days) and the median gestation of also 268 days.

To fit the normalised Gaussian function to this data, we centre the function at *d* = 270 days and choose the value of the standard deviation (SD) such that this function gives its best least-squares fit to the dataset. The standard deviation turned out to be SD = 9.89 days, and the resulting Gaussian fit is shown in Figure 1c on a $\log_{10}$ graph. Our model slope curve shown in Figure 1b is also plotted in Figure 1c on a $\log_{10}$ graph for comparison. Clearly, the gaussian and the model bell curves curve give almost identical results in the narrow region between *d* = 255 days and *d* = 285 days but rapidly diverge outside this region. The Gaussian function gives unrealistic results below d = 255 days and, therefore, underestimates the fraction of spontaneous births that are premature. Moreover, the Gaussian function is similarly unrealistic above *d* = 285 days since it does not reach zero at a finite value of d (it reaches zero only at infinity) and cannot predict a realistic maximum value for d. Thus, our model bell curve for the pregnancy delivery function fits the data better than the Gaussian function.

In the following discussion, the notation will be changed to suit the change in the dataset.

Using our realistic, saturated model function to fit SAT exam data

The *cumulative* fraction of an exam population that earns a fractional score of S *or less*, to be called the *performance* curve, will be denoted by P(S) so that P(S) varies between 0 and 1. Think of P(S) as a gauge of the performance of the cohort. In general, no one taking an ideal intelligence quota (I.Q.) exam will score 0 or 100% (S=0 or S=1) so that a true measure of the capability of a population can be ascertained. Thus, the mathematical properties of an ideal fractional performance function P(S) are

$$(2) \qquad P(0) = 0,\ dP(0)/dS = 0 \quad \text{and} \quad P(1) = 0,\ dP(1)/dS = 0$$

The distribution function PD(d) in Equation 1 satisfies the boundary conditions in Equation 2 by design.

Therefore, if the score on such an exam varies between 0 and 100% (*S* = 1), then an appropriate fractional performance distribution function that fits the boundary conditions in Equation 2 is:

$$(3) \qquad P(S) = [1 - exp(-\frac{S}{c(1-S)})]^{n}$$

where n is a positive number greater than 1, and c is a positive number greater than 0 that will be called the *critical* parameter. The values of n (called the *sharpness* parameter) and c are determined by a least-squares fit to actual exam data. The function in Equation 3 is almost identical to the one we used in Equation 1 to fit human gestation data.

Fr example, the scoring on the Scholastic Aptitude Exam (SAT) exam varies between 200 and 800, so the connection between the SAT(S) scores and the fractional score S is given by the equation

$$(4) \qquad SAT(S) = 200 + 800*S,$$

where SAT(0) = 200 and SAT(1) = 800. Because a small fraction of an SAT exam population gets a perfect score of 800, the function in Equation 3
must be modified because it does not satisfy the second boundary condition in Equation 2. So, SAT exam data will be modelled by using the modified function

$$(5) \qquad P(S) = [1 - exp(-\frac{S}{C(Sm-S)})]^{n},$$

where Sm is a 3rd added parameter denoting the *maximum* value of S (now greater than 1). We are deliberately allowing the top score on the SAT exam to increase beyond 800 to reach the *minimum* value *no one can attain* so that all the boundary conditions in (2) are satisfied. This will allow us to estimate the percentage of a population that achieves the perfect score of 800 on the exam. The function in Equation 5 is now identical to the one we used in Equation 1 to fit the human gestation dataset.

The derivative of the performance curve P(S) in Equation 5, dP(S)/dS, gives the probability distribution curve, our model bell curve. The function in Equation 5 will be shown to give excellent fits to *all* SAT data and will produce a bell curve that is a more accurate fit to the data than the Gaussian function.

We will begin the analysis of the 2021 SAT data by considering the performance of the National Reference Population (NRP) on the math exam. The math exam results for this population are plotted in Figure 2a along with the excellent fit to this data using Equation 5. Once the fractional performance function is ascertained, everything that can be asked about the results can be answered. For example, the score probability distribution curve, dP(S)/dS, can be immediately obtained by calculating the derivative of the fit function in Figure 2a, a result that is plotted in Figure 2b. In Figure 2b, we have our model bell curve, our replacement for the Gaussian distribution curve, and these points will be used to fit the Gaussian function to the same data. The peak in the model bell curve turns out to be at the SAT score of 500, so the most probable score is S_{mp} = 500. We propose breaking up the range of SAT scores into 7 performance categories as shown in Table 1 below.

Table 1. Percentage of Each Population in 7 Math Performance Categories on the 2021 SAT Exam

Math Distribution Population	Peak SAT score	<300 Very Poor	300-399 Poor	400-499 Low Average	500-599 High Average	600-699 Very Good	700-799 Excellent	800 Superior
National Reference	500	0.9%	12.9%	29.9%	30.1%	17.8%	6.7%	1.7%
Asian	675	0.3%	2.9%	9.6%	21.4%	33.4%	27.9%	4.5%
White	550	0.3%	6.3%	23.9%	36.5%	25.8%	6.9%	0.3%
Hispanic	450	1.7%	19.7 %	37.2%	28.3%	10.9%	2.1%	0.1%
African American	425	2.8%	26.7%	39.8%	23.2%	6.5%	0.8%	0.02%

Notice that in this model approach to fitting SAT data, the percentage of a population that scores the perfect SAT score of 800 can be estimated, something the gaussian function cannot do. The parameters used for this fit are shown in the first line of Table 2.

Table 2. Fit parameters for 2021 Math SAT Cohort Populations

Math SAT Population	*n* (sharpness parameter)	*c* (critical parameter)	*Sm* (maximum value of S parameter)
National Reference Population	5.2344	0.16977	2.032
Asian	4.301	0.51097	1.3786
White	5.9712	0.29563	1.4924
Hispanic	5.0711	0.18046	1.6863
African American	5.4061	0.12528	1.999

To fit the normalised Gaussian function to the data points in Figure 2c, we centre the function at the SAT score of 500 and choose the value of the standard deviation (SD) such that this function gives its best least- squares fit. The resulting Gaussian fit is shown in Figure 2c along with the points in Figure 2b for comparison. The standard deviation turned out to be 101.3 SAT points, practically the same as the 100-point SAT score range used to separate the 7 performance scores in Table 1. Clearly, the Gaussian and model bell curves give almost identical results in the narrow region between SAT=425 and SAT=575 but rapidly diverge outside this region. The Gaussian function gives unrealistically *high* results below SAT=400 and unrealistically *low* results above SAT =6 00. Thus, the model bell curve fits the data better than the Gaussian function.

As can be seen, our model bell curve is not symmetric around the most probable score S_{mp} = 500, in contrast with the less realistic, *symmetric* Gaussian function. Therefore, we replace the standard deviations of the Gaussian bell curve with the seven performance categories shown in Table (1), and we need only calculate the percentage of the population with scores within each category directly from the performance function P(S). The mathematics results of this analysis for the National Reference Population are shown in the first line of Table 1.

Similar performance curves can be calculated for the Asian, White, Hispanic, and African American datasets. For example, fitting the model function in Equation 5 to the African American (AA) population data is shown in Figure 3. The parameters used for this fit are shown in Table 2. Similar fits for the Hispanic, White, and Asian populations are obtained, and their respective fit parameters are also shown in Table 2. The performance fits for all four of these populations are shown in Figure 4a. Breakdowns of these results into our seven performance categories are also shown in Table 1.

The model bell curves associated with the African American, Hispanic, White, and Asian populations are shown in Figure 4b. The peaks in these bell curves occur at the SAT scores of 430 (African Americans), 460 (Hispanics), 560 (Whites), and 690 (Asians).

The percentage of each population scoring *below* the SAT score of 500, the average for the National Reference Population is 69.4% (African American), 58.7% (Hispanics), 30.5% (Whites), and 12.8% (Asians).

By contrast, the percentage of each population that scores *above* the SAT score of 700 is 0.8% (African Americans), 2.0% (Hispanics), 7.2% (Whites), and 33.4% (Asians), as seen in Table 2. Indeed, our estimate of the percentage of each population scoring the perfect score of 800 on the math SAT exam is 0.02% (African Americans), 0.1% (Hispanics), 0.3% (Whites), and 4.5% (Asians), as also shown in Table 2.

Mathematics is a gateway subject for succeeding in the science, technology, engineering, and mathematics (STEM) fields. In fact, African American and Hispanic college graduates are under-represented among degree recipients in STEM fields while Whites and Asians are over-represented. Unsurprisingly, therefore, African Americans and Hispanic workers are under-represented in the STEM workforce compared with their share of all workers, while Whites and Asians are over-represented. These facts suggest that Math SAT scores correlate with subsequent success in obtaining and holding STEM degrees and jobs.

For example, according to the American Institute of Physics analysis published in February 2021, in the two-year period of 2018 and 2019, out of a total of 1,910 Physics PhDs produced by our graduate schools, 42 (2.2%) were Hispanic-Americans, only 9 (0.47%) were African-Americans, 92 (4.8%) were Asian-Americans, 860 (45%) were white, and 887 (46%) were non-US citizens.

With figures like these, it is clear why the percentages of African-American (12.3% of the population) and Hispanic-Americans (12.3% of the population) PhDs in physics in our workforce are drastically under-represented. If you do not have a PhD in physics, you simply cannot apply for a position that requires it! To make this point unmistakably clear, it is not the fault of American companies that they employ so few blacks and Hispanic PhDs in physics - there simply are not enough of such PhDs to consider hiring.

Summary

Fitting the realistically saturated performance function P(S) in Equation 5 to distribution data and then deriving the model bell curve by computing the slope dP(S)/dS curve gives a better representation to datasets than does the gaussian bell curve.

References

College Board (2021). SAT Understanding Scores 2021

Jukic A. M., Baird D. D., Weinberg C.R., & McConnaughey D.R. Wilcox A.J.(2013). Length of human pregnancy and contributions to its natural variation. *Human Reproduction, 28* (1), 2848–2855.

Kennedy, B., Fry, R., & Funk, C. (2021). 6 facts about America's STEM workforce and those training for it. Pew Research Center.

Wolfram, MathWorld.

Fig. (1a). Fit to pregnancy duration data yielding Fractional PD(d) curve

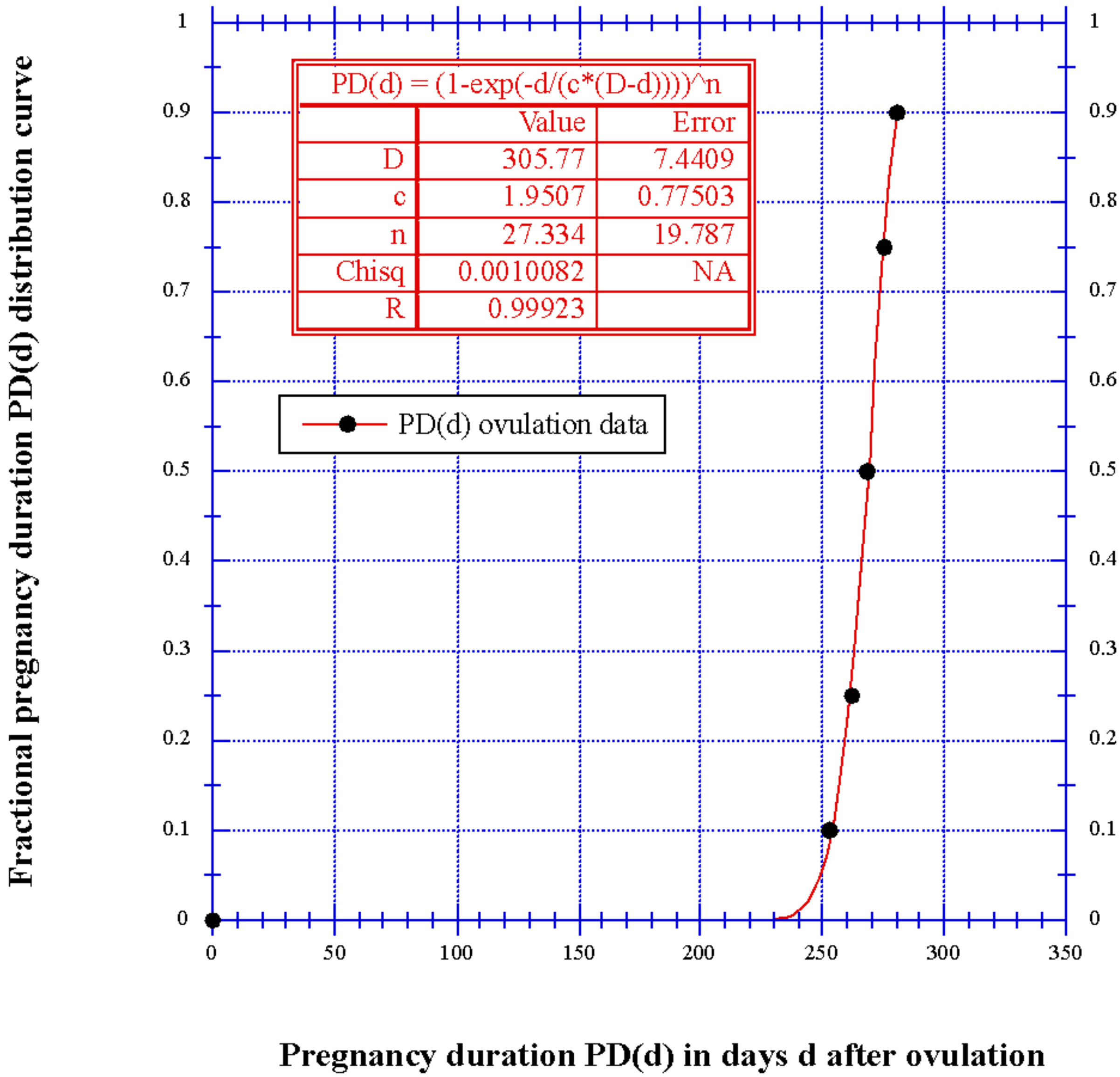

Fig. (1b). Slope of Fractional Pregnancy Duration Curve PD(d): the Model Fit Bell Curve

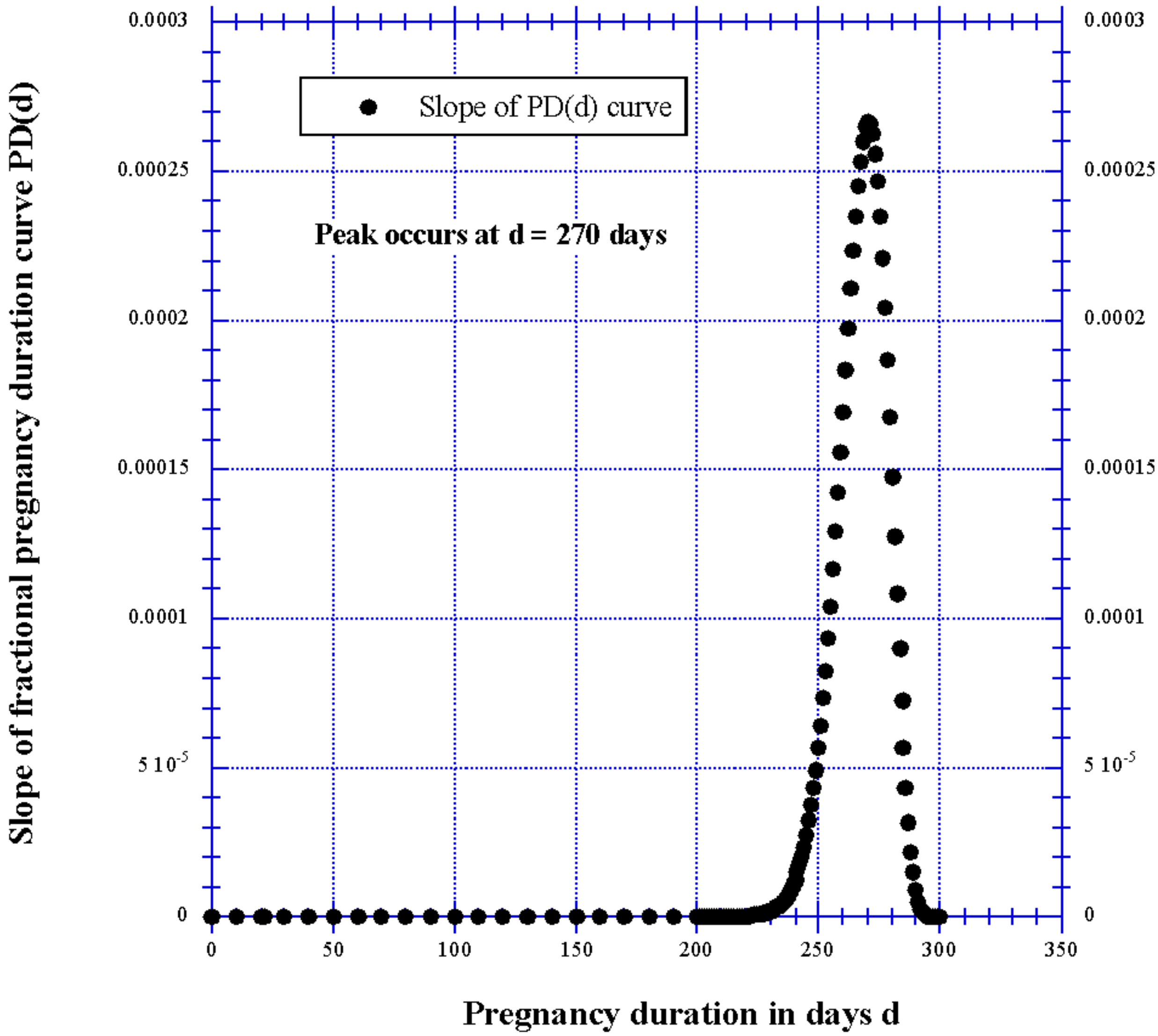

Fig. (1c). Fit of Gaussian function G(d) to slope of model pregnancy duration function PD(d) fit to data

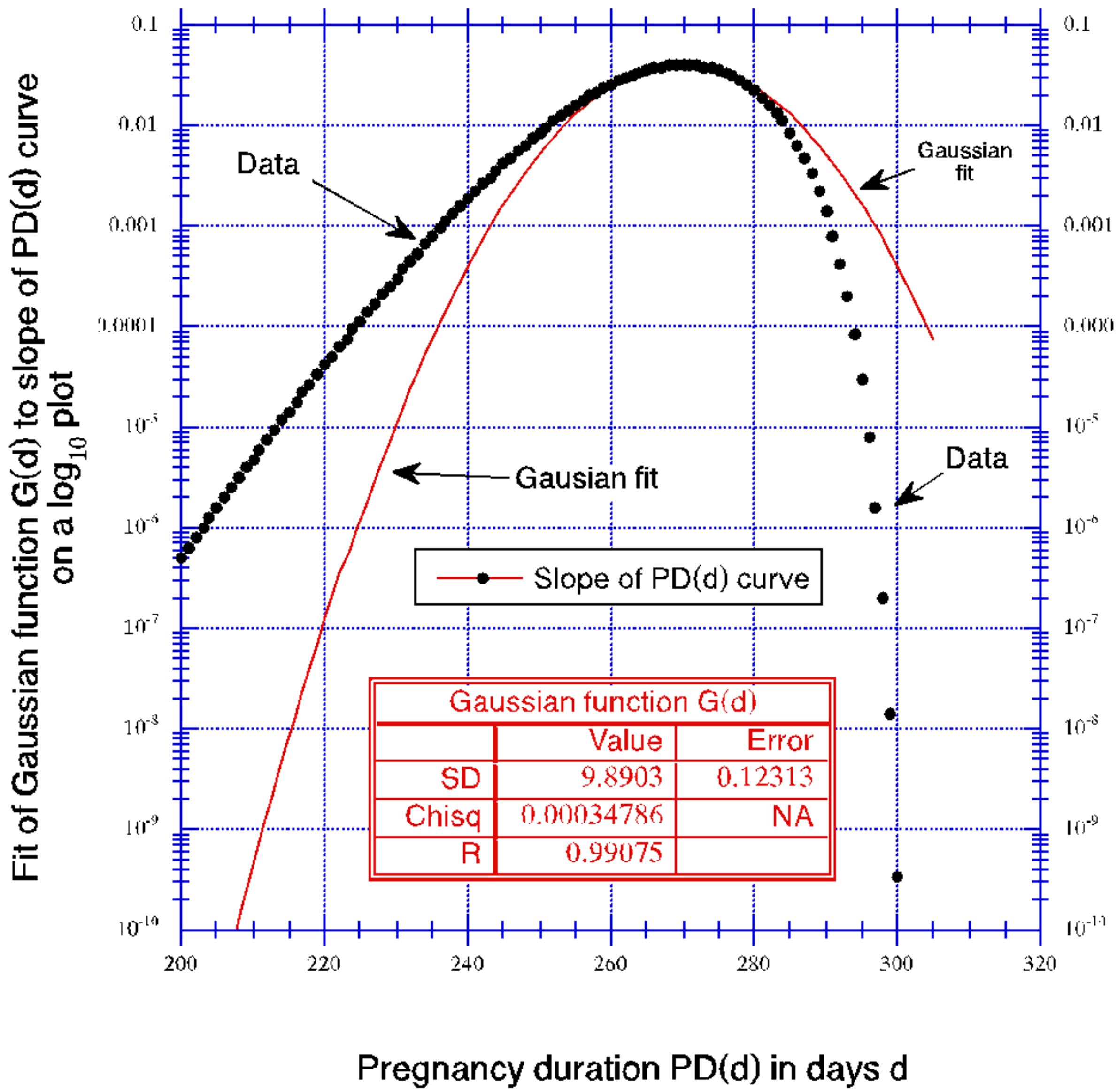

Fig (2a). Performance of REFERENCE population P(S) taking 2021 MATH SAT exam as a function of score S

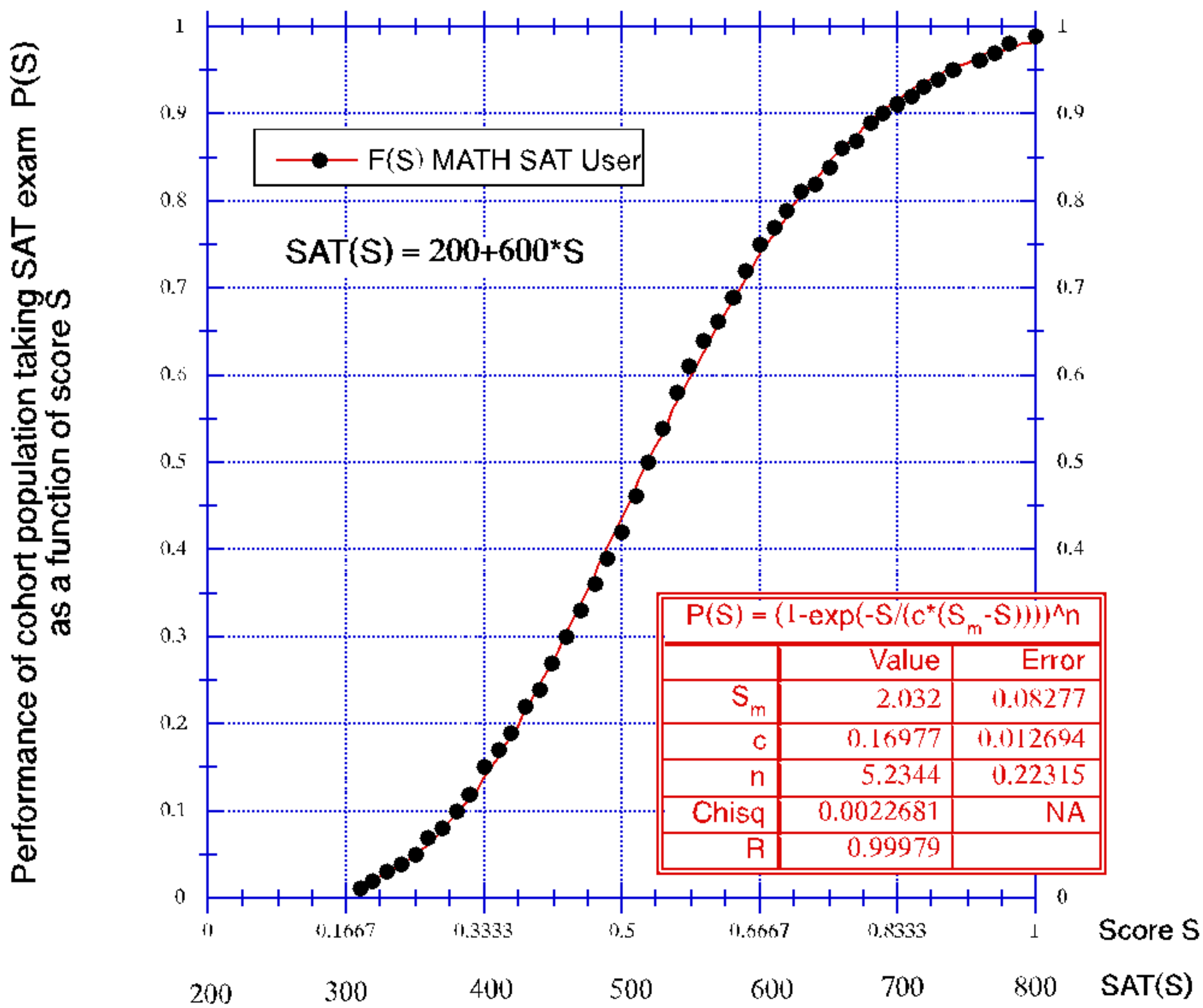

Fig. (2b). Slope of Model Fit to SAT Performance Data for Reference Population: the model Bell Curve, dP(S)/dS

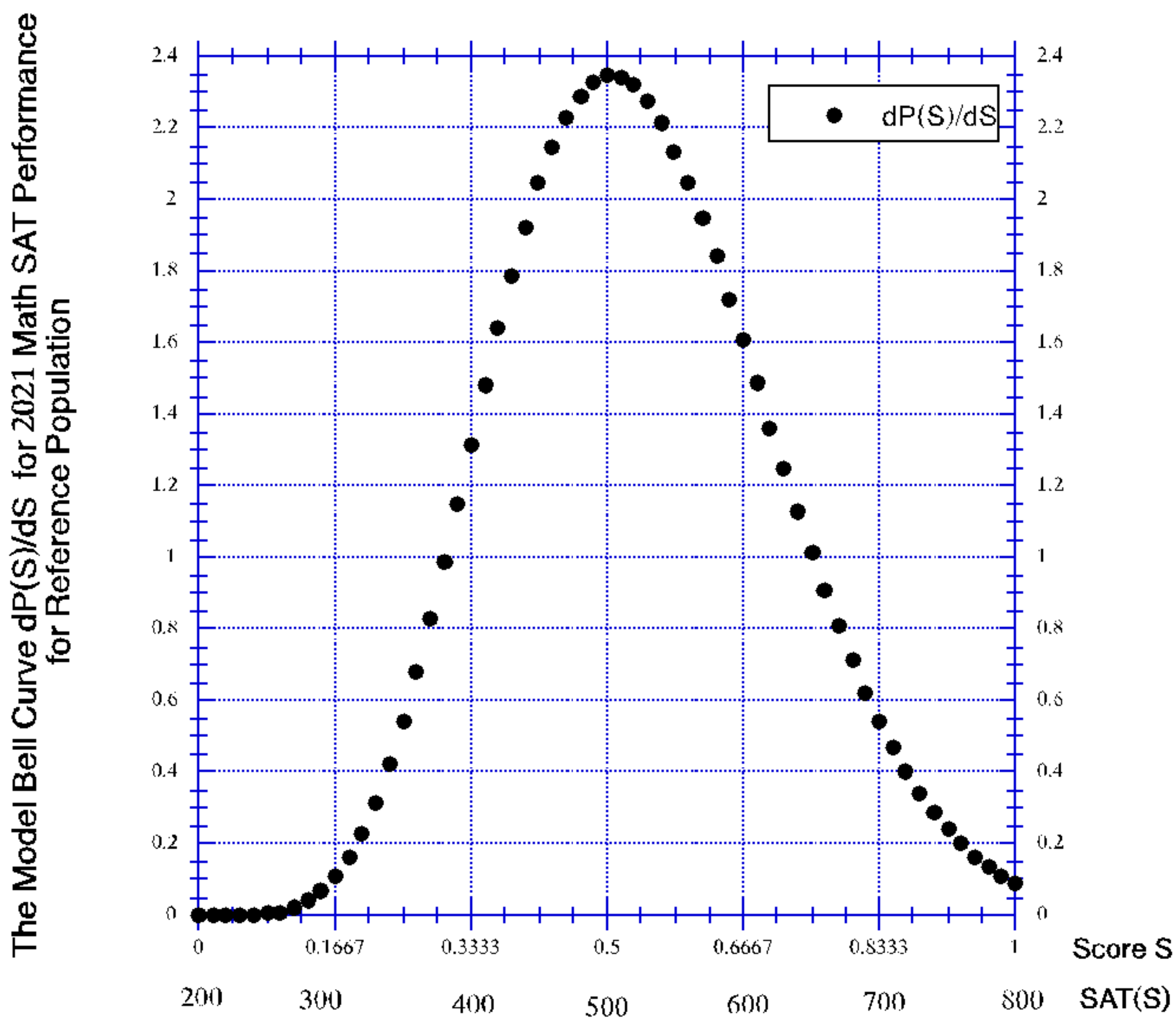

Fig. (2c). Gaussian function fit to dP(S)/dS curve fitted to Math SAT Data for National Reference Population

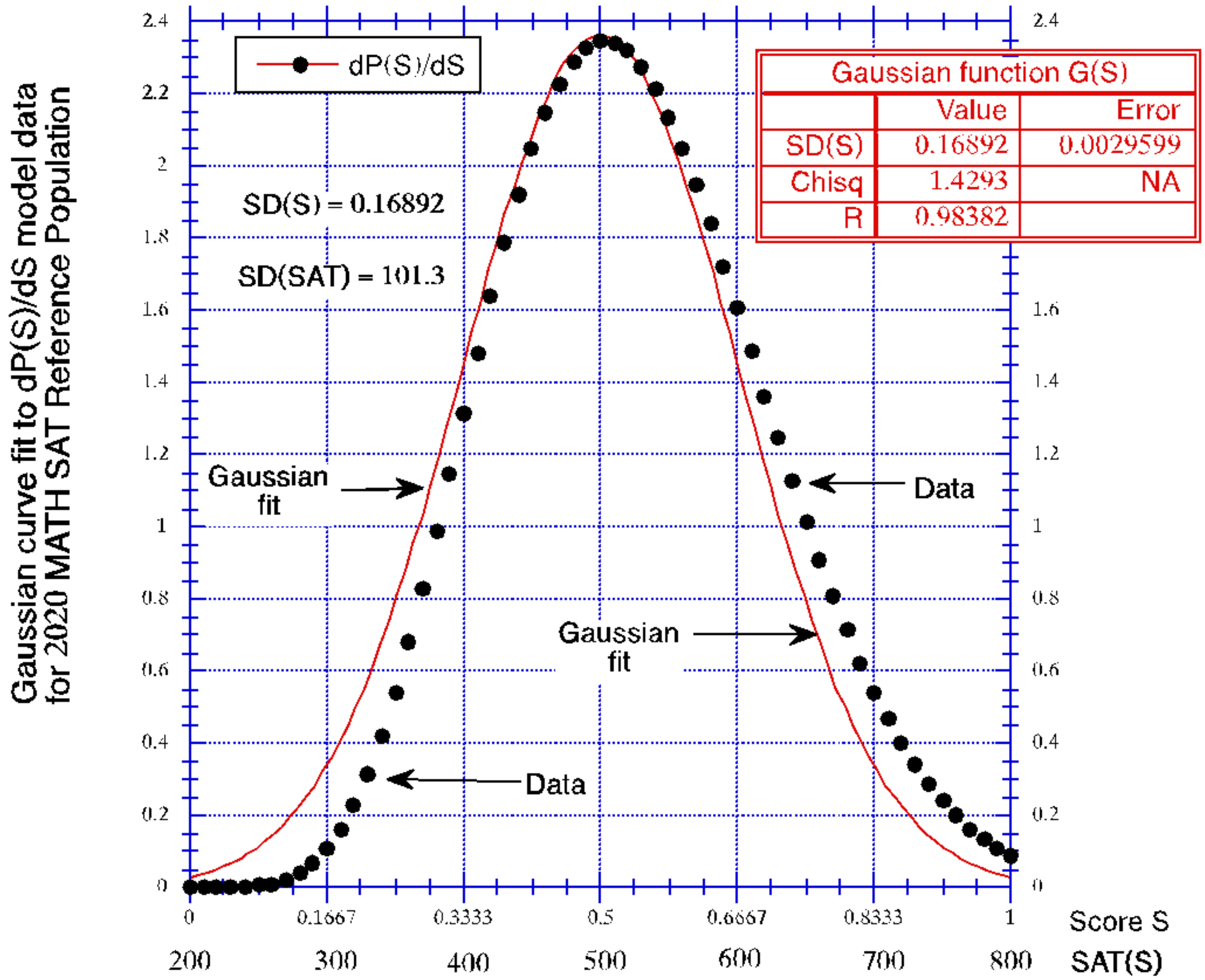

Fig (3). Performance of African American population P(S) taking 2021 Math SAT exam as a function of score S

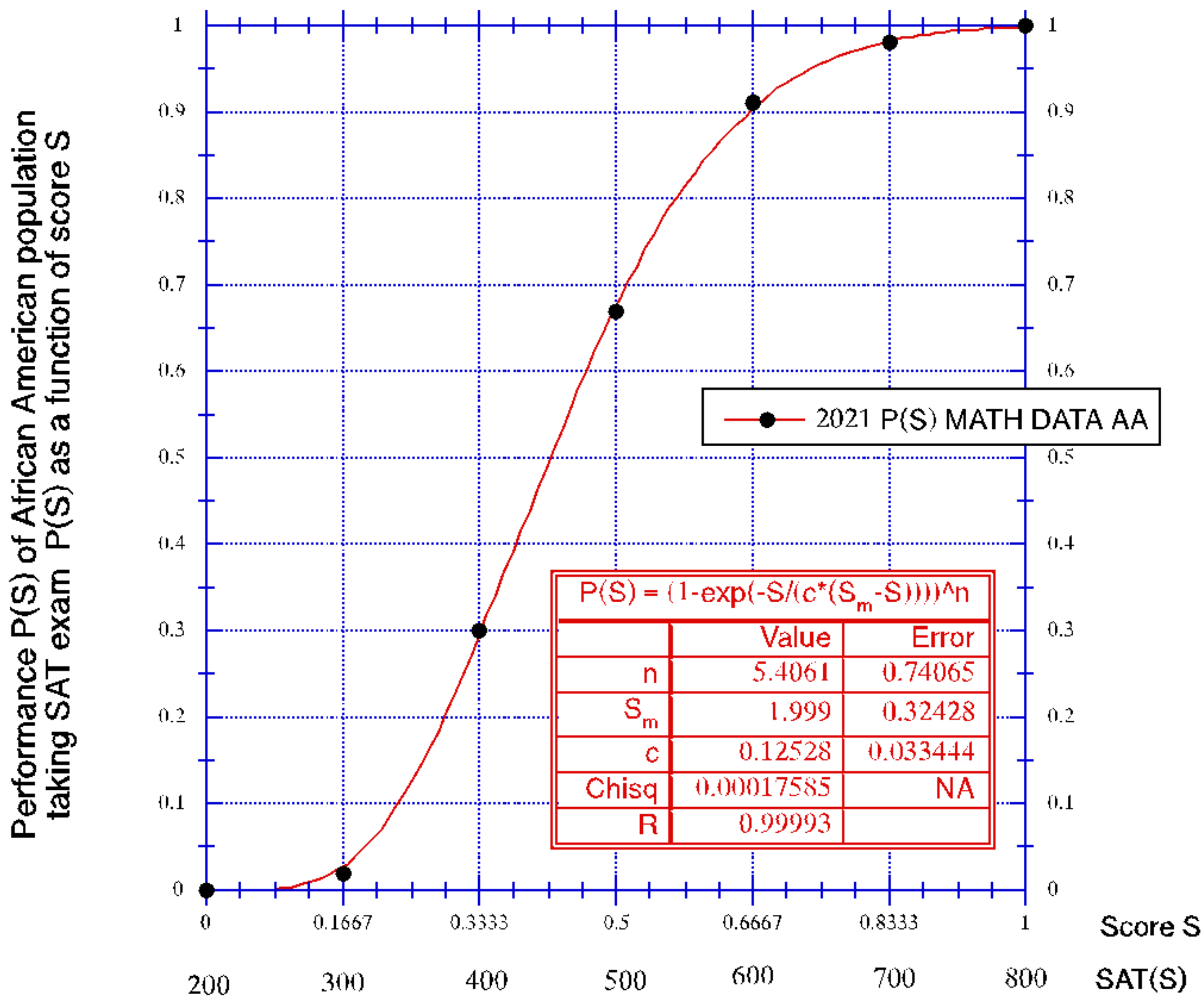

Fig. (4a). Math Performance curves P(S) for
Asian, White, Hispanic, and Black Populations

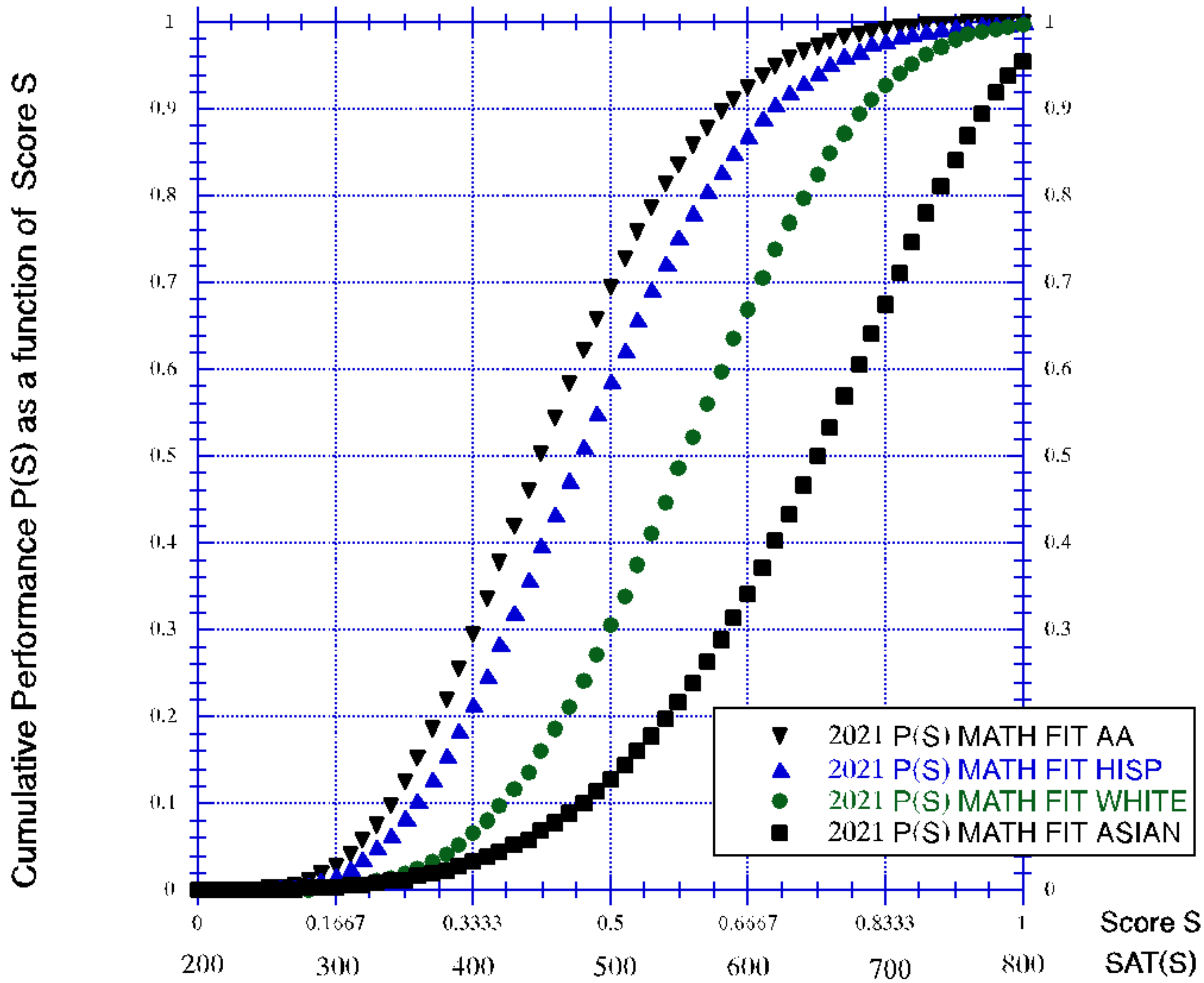

Fig. 4b. Model Fit Math Bell Curve Distributions, dP(S)/dS, for African American, Hispanic, White, and Asian Populations