1 Introduction

The USAPhO exam is the second exam in the selection process for the US Physics Team, which is often seen as the pinnacle of achievement in high school physics. It's hard to judge which scores correspond to which level of achievement (none, honorable mention, bronze, silver, gold, camp) since the actual scores of the exam are never released to students and hard cutoffs are never published.

However, in 2017, some statistics about the exam were published at the end of the 2017 USAPhO's solutions manual. They form the basis of our report, along with data from the medal list, qualifier list, and camper list.

2 Log-Normal Distributions

To understand the further analysis, one must first understand log-normal distributions. Consider a random variable X. In a normal distribution, we say that X is normally distributed. On the other hand, for a log-normal distribution, we say that $\ln(X)$ is normally distributed.

If X has a mean of μ_X and standard deviation σ_x, then its log-normal distribution will have

$$\mu = \ln\left(\frac{\mu_x^2}{\sqrt{\mu_x^2+\sigma_x^2}}\right)$$
$$\sigma = \sqrt{\ln\left(1+\frac{\sigma_x^2}{\mu_x^2}\right)}$$

Furthermore, its first quantile, median, and second quantile will be

$$Q_1 = \exp\left(\mu - \sigma\sqrt{2}\,\mathrm{erf}^{-1}\left(\frac{1}{2}\right)\right)$$
$$\tilde{x} = \exp\left(\mu\right)$$
$$Q_3 = \exp\left(\mu + \sigma\sqrt{2}\,\mathrm{erf}^{-1}\left(\frac{1}{2}\right)\right)$$

Here, $\mathrm{erf}^{-1} x$ is the inverse error function.

3 Data

Below are the parameters provided by AAPT at the end of the 2017 USAPhO Solutions.

Parameter	**Value**
Mean	53
Standard Deviation	29
Maximum	163
Upper Quartile	73
Median	49
Lower Quartile	30
Minimum	5

Achievement Level	**Count of People**
Qualifiers	487
Honorable Mention	93
Bronze Medal	89
Silver Medal	64
Gold Medal	43
Camper	22

4 Calculation of Parameters

Using $\mu_x = 53$ and $\sigma_x = 29$, we can find μ and σ.

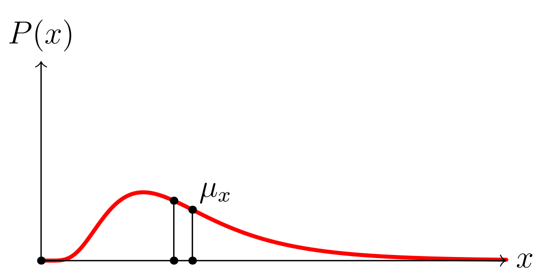
$$\mu = 3.839$$
$$\sigma = 0.512$$

5 Distribution

The probability density function (PDF) of the log-normal distribution is

$$P(X = x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$$

For the $\mu = 3.839$ and $\sigma = 0.262$ we found earlier, we can graph the distribution.



In order to validate this distribution, we can estimate some parameters using the formulae we have defined earlier.

Parameter	Actual Parameter Value	Calculated Parameter Value
Mean	53	53
Standard Deviation	29	29
Upper Quartile	73	65.650
Median	49	46.479
Lower Quartile	30	32.906

We can therefore see that the distribution is fairly accurate for lower scores, and starts diverging for higher scores, as evidenced by the difference in the upper quartile values. We proceed, then, with caution, as much of this estimation is based on that.

6 Score Estimation

The USAPhO scores are often determined in percentiles, so it's useful to calculate this. Let a student score above a percent p of USAPhO takers. We want to find what score they get (in statistics this is known as the quantile function). To do this, we use the CDF of the log-normal distribution.

$$F(x) = \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left(\frac{\ln(x) - \mu}{\sigma\sqrt{2}}\right)$$

Setting this equal to p, we solve and find that the score x needed to score above p percent of testakers is

$$x = \exp(\mu + \sigma\sqrt{2}\,\mathrm{erf}^{-1}(2p - 1))$$

We can finally use the data we've gotten from the medal list, but must first tabulate it into a more approachable form. A key assumption we must make here is the number of people that take the USAPhO after qualifying for it. Let the number of people be n.