Theory: Bayesian Reasoning, Aggregate Paradoxes and Power Law

1. The following table shows two basketball players named *Really Bad* and *Totally Inconsistent*, and their free throw averages for seasons 2018 and 2019 respectively.

	Really Bad	Totally Inconsistent
2018	30%	90%
2019	20%	25%

How many free throws each of these two players had to make each year in order for Really Bad to have a better overall percentage (for both 2018 and 2019 seasons) free throws than Totally Inconsistent? Of course there could be many answers – pick one of them and justify.

In other words – specify how many free throws Really Bad had to throw in 2018, 2019 and how many did Totally Inconsistent had to throw
In 2018 and 2019 – in order to have lower overall free throw percentage than Really Bad.

Ans: When players have a difference in their weightage their overall percentage could vary. So if Really Bad had 500 free throws in 2018, while Totally Inconsistent had only 10 free throws then Really bad got it into the basket 150 times while Totally Inconsistent got it in only 9 times.

In 2019 if Really Bad had 10 free throws then he got 2 free throws in. In 2019 if Totally Inconsistent had 10000 throws then he got only 2500 free throws in.

So overall Really Bad got (150 + 2) / 510 = 29.8% of his throws in. While, Totally Inconsistent got (9 + 2500) / 10010 = 25.1% of his throws in.

2. Biden won 80% of the vote in one of “Blue” states and Trump the remaining 20%. There are 100 counties in that state.

Which of the following are true (could be more than one)

A. Biden definitely won at least 80 counties (i:e: got more votes than Trump in these counties)

B. Biden definitely won at least 20 counties (i:e: got more votes than Trump in these counties)

C. It is possible that Trump won 100% vote in 99 counties

For first two options if true, prove it. For last two options if true - show how it is possible.

Ans: The last option(C) is correct because if the 99 counties that trump won in had a population of just 1 while the 100th county had a population of 396, where Biden won 100% of the votes, then Biden clearly won in that state because he had more votes overall.

3. A paper has several typos on the first page. What is the probability that the paper will be rejected for a conference? The ratio of rejection is 80%. The probability that rejected papers have typos on the first page is 40%, and the probability that accepted paper have typos on the first page is 10%.

Ans: Prior odds = 8:2 = 4:1
True Positive = 40% = 0.4
False Positive = 0.1
Likelihood ratio = 0.4 / 0.1 = 4
Posterior odds = 4 * (4 / 1) = 16:1 = 16 / 17 = 0.941

4. At an interview a candidate makes a bad joke. What are the odds that he won't be hired? The odds for Company hiring interview candidates to be hired are 1:1. The hired ones make a bad joke 5% of the time, while the rejected candidates, 10% of the time.

Ans: Prior odds = 1:1
True Positive = 10% = 0.1
False Positive = 5% = 0.05
Likelihood ratio = 0.1 / 0.05 = 2
Posterior odds = 2 * (1 / 1) = 2:1 = 2 / 3 = .667

5) Lets assume that Book sales are distributed according to the power law with exponent = 2. Suppose 1000 books sold over 100,000 copies. How many books are expected to sell more than 1 million copies?

Ans: 1 million copies is 10 times 100,000 copies. Since, 1000 books sell over 100,000 copies, then the number of books would decrease by 10 ^ exponent which is 10^2 = 100. Hence, the number of books that would sell over 1 million copies is 1000 / 100, which is 10.

6) What if books sales were distributed according to normal distribution with average sales of 10000 books and standard deviation of 10,000. If there was total of 1 million books published, how many books would be expected to sell more than 50,000 copies?

Ans: The z value would be (50,000 - 10,000) / 10,000 which is 4. 1 Standard deviation of the mean consists 68% of the books, 2 standard deviation contains 95% of the books, and 3 standard deviations consist of 99.7% of the book. The normal distributed curve is very steep, so anything after the 4th

standard deviation is so tiny (which also has to be divided by 2 since it's on both sides of the curve) that it's almost negligible. Hence, the number of books that sold over 50,000 copies is 0.