MATH 408 Homework 1

Qilin Ye

August 24, 2021

Problem 3

Two people take turns throwing darts at a board. Person A goes first, and each of their throws has a probability of $1/4$ of hitting the bullseye. Person B goes next, and each of their throws has a probability of $1/3$ of hitting the bullseye. Then Person A goes, and so on. With what probability will Person A hit the bullseye before Person B does?

Solution. Let S_n be the event in which the n^{th} shot is the first to hit the bullseye. (For example, in S_4, A misses, then B misses, then A misses again, and finally B hits.) It is clear that the $S_n's$ are pairwise disjoint and that the events in which A hits the bullseye before B does is

$$S := \bigcup_{n=1}^{\infty} S_{2n-1}.$$

Therefore,

$$\begin{aligned} P(S) &= \sum_{n=1}^{\infty} P(S_{2n-1}) \\ &= \frac{1}{4} + \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{4} + \left(\frac{3}{4} \cdot \frac{2}{3}\right)^2 \cdot \frac{1}{4} + \left(\frac{3}{4} \cdot \frac{2}{3}\right)^3 \cdot \frac{1}{4} + \dots \\ &= \frac{1}{4} \cdot \left(1 + \frac{1}{2} + \frac{1}{4} + \dots\right) = \frac{1}{2}. \end{aligned}$$

Problem 4

Suppose you have a car with twenty tires, and the car mechanic removes all twenty tires. Suppose the mechanic now puts the tires back on randomly, so that all arrangements of the tires are equally likely. With what probability will no tire end up in its original position? Give an answer to ten decimal places of accuracy (e.g. your answer could be 0.1234567891). Can you guarantee that these ten decimal places are correct?

Solution. It is clear that there are $20!$ (factorial) possible arrangements. We are interested in the number of derangements. Since 20 is not very small, it might be more convenient to directly compute the derangement number $!20$ using the recursion formula

$$!n = (n-1)(!(n-1) + !(n-2)) \qquad \text{subject to } !0 = 1 \text{ and } !1 = 0.$$

Upon some simple calculations, we obtain the answer

$$\frac{!20}{20!} \approx 0.3678794412.$$

Of course, we could also compute the number of cases in which at least one tire returns to its original position using inclusion-exclusion, but since 20 is fairly large, this might require more work computationally.

Problem 5

Suppose a test for a disease is 99.9% accurate. That is, if you have the disease, the test will be positive with 99.9% probability. And if you do not have the disease, the test will be negative with 99.9% probability. Suppose also the disease is fairly rare, so that roughly 1 in 20,000 people have the disease. If you test positive for the disease, with what probability do you actually have the disease?

Solution. We apply Bayes' Theorem.

	Disease (D)	No Disease (ND)
Positive (+)	$0.999 \cdot 1/20000$	$0.001 \cdot 19999/20000$
Negative (-)	$0.001 \cdot 1/20000$	$0.999 \cdot 19999/20000$

$$P(D|+) = \frac{P(+|D)P(D)}{P(+)} = \frac{0.999/20000}{0.999/20000 + 0.001 \cdot 19999/20000} \approx 0.048.$$

Problem 6

Suppose I tell you that the following list of 20 numbers is a random sample from a Gaussian random variable, but I dont tell you the mean or standard deviation.

5.1715, 3.2925, 5.2172, 6.1302, 4.9889, 5.5347, 5.2269, 4.1966, 4.7939, 3.7127
5.3884, 3.3529, 3.4311, 3.6905, 1.5557, 5.9384, 4.8252, 3.7451, 5.8703, 2.7885

To the best of your ability, determine what the mean and standard deviation are of this random variable. (This question is a bit open-ended, so there could be more than one correct way of justifying your answer.)

Solution. For convenience denote the above numbers using $X_1, ..., X_{20}$. The most reasonable estimation of the sample mean of these 20 numbers is simply

$$\overline{X} = \frac{1}{20}\sum_{i=1}^{20} X_i \approx 4.4426.$$

The best estimator for sample variance is given by

$$S^2 = \frac{1}{20-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \approx 1.4618.$$

Taking square root, we obtain the standard deviation 1.2091. □

Problem 7

Suppose I tell you that the following list of 20 numbers is a random sample from a Gaussian random variable but I don't tell you the mean or standard deviation. Also, around one or two of the numbers was corrupted by error so that it is totally unrelated to the actual Gaussian random variable.

−1.2045	−1.4829	−0.3616	−0.3743	−2.7298	−1.0601	−1.3298	0.2554	6.1865	1.2185
−2.7273	−0.8453	−3.4282	−3.2270	−1.0137	2.0653	−5.5393	−0.2572	−1.4512	1.2347

To the best of your ability, determine what the mean and standard deviation are of this random variable. Supposing you had instead a billion numbers and about $5 - 10$ percent of them were corrupted samples, can you come up with some automatic way of throwing out the corrupted samples?

Solution. The sample mean of these data is approximately $\overline{X} = -0.8036$. Computing each $(x_i - \overline{X})^2$ yields the following table:

x_i	$(x_i - \overline{X})^2$	x_i	$(x_i - \overline{X})^2$	x_i	$(x_i - \overline{X})^2$	x_i	$(x_i - \overline{X})^2$
−1.2045	0.1607	−1.4829	0.4615	−0.3616	0.1954	−0.3743	0.1843
−2.7298	3.7103	−1.0601	0.0658	−1.3298	0.2769	0.2554	1.1215
6.1865	48.8614	1.2185	4.0888	−2.7273	3.7007	−0.8453	0.0017
−3.4282	6.8886	−3.2270	5.8729	−1.0137	0.0441	2.0653	8.2305
−5.5393	22.4269	−0.2572	−0.2985	−1.4512	0.4194	1.2347	4.1546

We clearly notice that 6.1865 and −5.5393 behave abnormally as their corresponding $(x_i - \overline{X})^2$ are much larger than all others. After discarding these two and using the formula in the previous problem, we obtain sample mean −0.9288 and standard deviation 1.5261.

I am not sure how to generalize this into the case of a billion numbers. Perhaps we should discard some numbers so that the $68 - 95 - 99.7$ rule roughly applies?