Two Pens In a Pocket Must Be Different: A Nerd-Oriented Lesson From Statistics

Olga Kosheleva[1] and Vladik Kreinovich[2]
[1]Department of Teacher Education
[2]Department of Computer Science
University of Texas at El Paso
500 W, University
El Paso, TX 79968, USA
olgak@utep.edu, vladik@utep.edu

Abstract

Some people always carry a pen with them, so that if an idea comes to mind, they will always be able to write it down. Pens sometimes run out of ink. So, just in case, people carry two pens. The problem is that often, when one carries two identical pens, they seem to run out of ink at about the same time – which defeats the whole purpose of carrying two pens. In this paper, we provide a simple statistics-based explanation of this phenomenon, and show that a seemingly natural idea of carrying three pens will not help. The only way to avoid the situation when all the pens stop working at about the same time is to carry pens of different types.

1 Puzzling Situation

Carrying pens: a general idea. Sometimes, an idea come to mind when you are walking or hiking or taking a bus. If you do not write it down, you may forget it.

So, some people carry a pen in their pocket – if an idea comes, they will be able to write it down.

Pens sometimes run out of ink. So, to guarantee that there will be a working pen available, a good idea is to carry two pens.

What happened. A colleague bought a dozen of identical pens, and selected two of them to carry. After a while, as expected, one of them stopped working. Somewhat unexpectedly, shortly after, the second one stopped working, so this person was left without a pen to write.

How come? Shall this person carry three pens next time? Will that help?

This can be a good learning exercise in statistics. After a while, this person understood what happened – and realized that this can be a good pedagogical exercise in statistics.

2 Solution to the Puzzle

Analysis of the problem. Since the pens are identical, each time the person needs to write down something, this person selects a pen at random, with equal probability $\frac{1}{2}$ of selecting each of the two pens. Suppose that overall, this person made N notes.

How many notes n were made by the first pen? To find this number, let us introduce, for each of the notes $i = 1, \ldots, N$, the value x_i which is equal to 1 if the first pen was used for the i-th note and 0 if the second pen was used. Then, n is equal to the sum of all these values:

$$n = \sum_{i=1}^{n} x_i. \tag{1}$$

Thus, n is the sum of the large number of independent random variables. The value N is large: pens are good, and notes are usually short. Thus, by the Central Limit Theorem (see, e.g., [1]), the distribution of n should be close to normal. To be more precise, the Central Limit Theorem says that when N tends to infinity, the distribution of the sum of N variables tends to normal – which exactly means that if N is large, then the distribution of the sum is close to normal.

For each i, the mean value of the random variable x_i is equal to

$$E[x_i] = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0 = 1/2,$$

and its variance is equal to

$$V_i = E\left[(x_i - E[x_i])^2\right].$$

Here, for both possible values $x_i = 1$ and $x_i = 0$, we have

$$(x_i - E[x_i])^2 = \left(x_i - \frac{1}{2}\right)^2 = \frac{1}{4},$$

so $V_i = \frac{1}{4}$.

For the sum of several independent random variables, the mean is equal to the sum of the means, and the variance is equal to the sum of the variances. Thus, for the sum n of N random variables x_i, the mean is equal to $\mu = \frac{N}{2}$, and the variance is equal to $V = \frac{N}{4}$. Thus, the standard deviation is equal to

$$\sigma = \sqrt{V} = \frac{\sqrt{N}}{2}.$$

For normal distribution, with probability 95%, the value is within the two-sigma interval

$$[\mu - 2\sigma, \mu + 2\sigma] = \left[\frac{N}{2} - \sqrt{N}, \frac{N}{2} + \sqrt{N}\right].$$

Resulting estimate explains the puzzle. Suppose that the pens lasted for two months, and, on average, a person makes 10 notes a day. During this time, a person made $N = 60 \cdot 10 = 600$ notes. This, with high probability, the number of notes made by the first pen lies within the interval

$$\left[\frac{600}{2} - \sqrt{600}, \frac{600}{2} + \sqrt{600}\right] \approx [300 - 24, 300 + 24].$$

On average, a pen made about 300 notes and ran out of ink, so a pen lasts for about 300 notes. On average, each pen makes 5 notes a day, so:

- $300 - 24$ corresponds to $60 - 5$ days, and
- $300 + 24$ to $60 + 5$ days.

Thus, both pends run out of ink within 5 days of the 2-months moment.

Thus, when a person goes on a typical week-long trip – e.g., to an overseas conference – there is a high probability that if one pen runs out of ink, the second pen will run out of ink during the same trip. This may be not so bad if it is a conference – one can go to a store and buy a pen, and many conferences give out pens as part of the registration package – but what if it is a week-long hike?

3 What If We Take Three Pens?

Will it help if a person takes three pens?

Not really, the same arguments show that all three pends will run out of ink at the same time. Same with four pens, etc.

4 So What Can We Do?

There is no way to avoid this problem with identical pens. So, the only possible solution is to use *different* pens. Then:

- we can use one of the pens until it runs out of ink, and
- only then start using the second pen.

5 This Is More Serious Than About Pens

Duplication is not only for nerds who carry pens in their pockets. In many other situations, we have several devices, just in case on of them fails. And the pens example brings a lesson:

- it is *not* a good idea to use one of the devices at random, this defeats the purpose of reliability,
- it is better to have one of the devices idle.

This may be not the most efficient use of resources, but this is the only way to make a system more reliable.

Acknowledgments

This work was supported in part by the US National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science) and HRD-1242122 (Cyber-ShARE Center of Excellence).

References

[1] D. J. Sheskin, *Handbook of Parametric and Non-Parametric Statistical Procedures*, Chapman & Hall/CRC, London, UK, 2011.