

Can You Swim With the Fishes?



October 2026

A puzzle from Fiddler on the Proof, about a sensor that cannot help flattering the fast.

Problem

Two koi share a long straight pond. Nemo swims at one mile per hour and Dory at two, and both spend the day going back and forth. A sensor in the water records the speed of every fish that passes it. After several days of logging, what is the average of the recorded speeds?

Solution

The tempting answer is the average of the two speeds, 1.5 miles per hour. It is wrong, and the reason is worth more than the arithmetic.

The log does not contain one entry per fish. It contains one entry per passing, and the two fish do not pass at the same rate. A fish of speed v in a pond of length ℓ needs time ℓ/v to cross from one end to the other, and it goes past the sensor once on each such trip. Its entries therefore arrive at a rate of v/ℓ per hour, in direct proportion to its speed. Dory, swimming twice as fast, appears in the log twice as often.

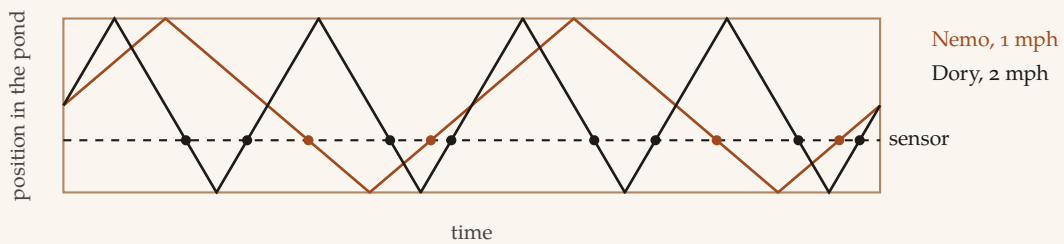


Figure 1: The two fish bouncing between the ends of the pond, with the sensor fixed at one place. Each dot is an entry in the log. Dory's path is steeper, so it meets the sensor line twice as often.

So in a long log, two entries in every three are Dory's and one is Nemo's, and the average recorded speed weights each speed by itself:

$$\frac{1 \cdot 1 + 2 \cdot 2}{1 + 2} = \frac{5}{3}.$$

$$\text{average recorded speed} = \frac{5}{3} \approx 1.667 \text{ miles per hour.}$$

The sensor is not faulty and the fish are not unusual. The sampling is simply not uniform over fish: a fish that moves more gets measured more. The same effect explains why a bus stop feels slower than the timetable says and why asking people how large their class is gives too large an answer. Every observation is made by something the observed object had to arrive at, and arriving is what the fast do more often.

Extra Credit

Now the pond holds many koi, all starting at the centre, with velocities drawn from a normal distribution of mean 0 and standard deviation 1 mile per hour. The sensor records the speed of every fish that passes it, with direction ignored. What value does the average recorded speed approach?

Weighting by speed, in general

The counting argument does not care how many fish there are or how their speeds are spread. A fish of speed s still passes the sensor at a rate proportional to s , so in a long log the chance that any given entry belongs to a fish of speed about s is proportional to s times how common that speed is among the fish. If the fish have speed density f , the log has density

$$g(s) = \frac{sf(s)}{\mathbb{E}[S]},$$

the denominator being what makes it a density. Averaging against g gives the one formula that runs the whole puzzle:

$$\text{average recorded speed} = \frac{\mathbb{E}[S^2]}{\mathbb{E}[S]}.$$

For the first part, with speeds 1 and 2 equally common, this is $(1 + 4)/(1 + 2) = 5/3$ again.

The normal case

Here the velocity V is standard normal and the speed is $S = |V|$. The two moments needed are familiar: $\mathbb{E}[S^2] = \mathbb{E}[V^2] = 1$, and $\mathbb{E}[S] = \sqrt{2/\pi}$, the mean of a half-normal. Hence

$$\text{average recorded speed} = \frac{1}{\sqrt{2/\pi}} = \sqrt{\frac{\pi}{2}} \approx 1.253 \text{ miles per hour.}$$

Two remarks are worth making. The first is that the answer is exactly $\pi/2$ times the average speed of the fish themselves, which is $\sqrt{2/\pi} = 0.798$. Weighting by speed inflates the figure by about 57 per cent, and that factor depends only on the shape of the distribution, not on its scale.

The second is what the log actually looks like. Multiplying the half-normal density $\sqrt{2/\pi} e^{-s^2/2}$ by s and normalising gives

$$g(s) = s e^{-s^2/2}, \quad s > 0,$$

which is the Rayleigh distribution. The sensor's log is therefore distributed exactly as the speed of a particle whose velocity is normal in *two* dimensions rather than one, and the mean of that distribution is $\sqrt{\pi/2}$, as computed.

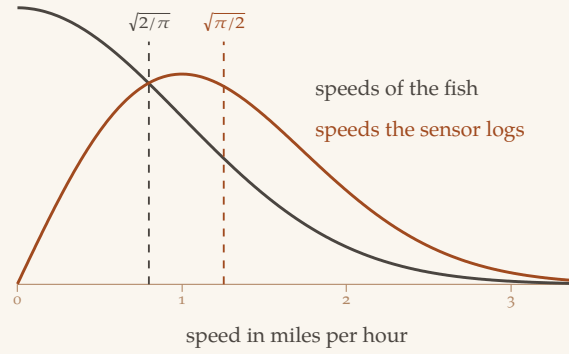


Figure 2: The speeds of the fish, a half-normal, against the speeds the sensor logs, a Rayleigh. Weighting by speed empties the log of the slow fish and moves the average from $\sqrt{2/\pi}$ to $\sqrt{\pi/2}$.

The computation

The check counts crossings rather than assuming the rate law that the argument rests on. A bouncing path is unfolded into a straight one, so a fish starting at x_0 with speed s is at the sensor whenever $x_0 + st$ is congruent to $\pm p$ modulo twice the pond's length. The unfolding is itself checked against a brute-force count of sign changes of the folded path.

```
def crossings_unfold(v, T):
    """times t in [0,T] with the unfolded path x0 + |v| t equal to +-P mod 2L"""
    s, hi, n = abs(v), X0 + abs(v)*T, 0
    for target in (P, -P):
        k_lo = np.ceil((X0 - target)/(2*L))
        k_hi = np.floor((hi - target)/(2*L))
        n += max(0, int(k_hi - k_lo) + 1)
    return n
```

The two agree exactly at every speed tried. Over a long run the fish of speed 1 and 2 produce 20,000 and 40,000 entries, a ratio of exactly 2, and the logged average is 1.666667.

For the extra credit, counting crossings for two hundred thousand fish gives a logged average of 1.2570, against $\mathbb{E}[S^2]/\mathbb{E}[S] = 1.2570$ for that same sample, which is what confirms the weighting. The remaining gap to $\sqrt{\pi/2}$ is the sampling error in the speeds themselves, and it closes with more fish: twenty million give 1.253286 against 1.253314, and the ratio of the logged mean to the true mean comes to 1.570880 against $\pi/2 = 1.570796$.