Victoria Bunimovich

Lane Inequality in Competitive Swimming: A Mathematical Investigation of Disqualifications

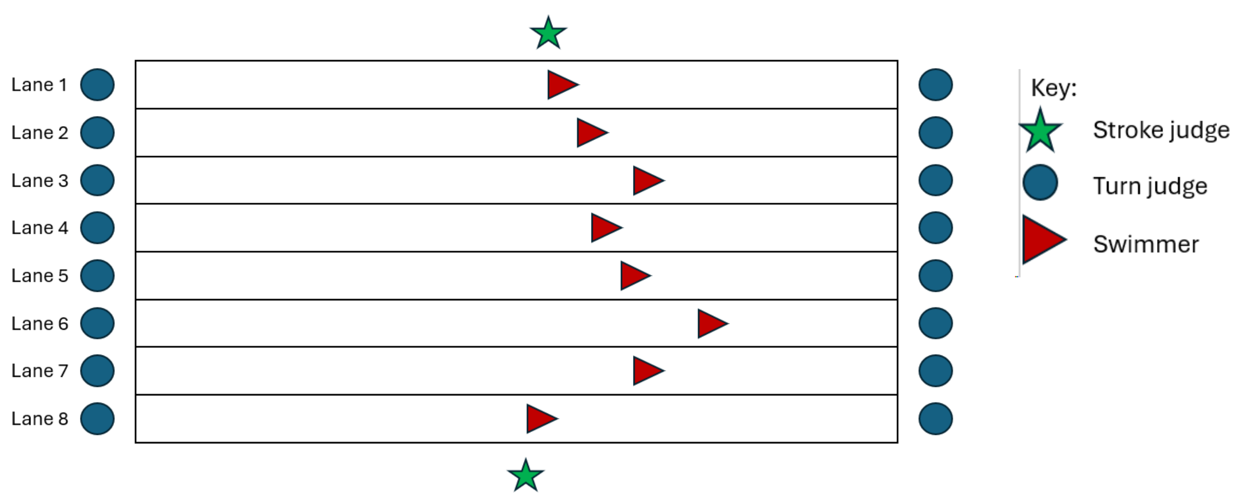
Have you ever watched a swimming race? Unless you are involved in competitive swimming like my sister and I, you may have watched the Olympics on TV. Everything looked so simple and smooth – all swimmers had almost identical movements, some marginally quicker than others. Have you noticed the judges? Probably not. It is also unlikely that you have seen anyone disqualified and if you have, it was probably for diving in too early.

A few weeks ago, my sister and I were racing at the County swimming championships. Not exactly the Olympics, but a decently advanced level of competition. After one of her races, my sister got disqualified for an alternating leg kick while swimming butterfly. She was not too upset as this was not her best race regardless. Jokingly, she told me, “This is all because I was in lane 8. Next time I am in an end lane, I won’t even bother jumping in as they will disqualify me anyway!”. After laughing with her, I went back to my team and found out that one of my teammates also got disqualified in a different race, but this time she swam in lane 1. This got me thinking. I’ve heard a few times that end lanes of the pool receive more scrutiny from judges, but until this time, I have not really asked myself to what extent this is the case.

A Brief Overview of Judging in Competitive Swimming

Let me quickly explain how swimming races are judged. In fact, there are many different rules that govern the way each stroke should be swum. For example, in butterfly, movements of arms and legs must be simultaneous, while in breaststroke, there must be one leg kick per arm stroke, and in freestyle, backstroke and butterfly, a swimmer is not allowed to swim underwater with limited exceptions. During the race, these rules are enforced by so-called stroke judges, unless a swimmer is performing a turn at the end of the pool, in which case they are enforced by turn judges. Stroke judges walk along the side of the pool back and forth to check for any breaches of stroke rules. If they see a swimmer, for example, swimming underwater in the middle of the pool during a freestyle race, they will disqualify them. There may be one or two stroke judges on each side.

A diagram to show the placement of stroke and turn judges:



It seems natural that stroke judges will see the lanes closest to them the best. The rules also say that if a judge is not completely sure whether an infraction has occurred, the swimmer always gets the benefit of the doubt. All decisions must be made independently, meaning that the judges are not allowed to discuss among themselves whether an infraction has occurred.

Potential Bias towards the End Lanes

It would then seem natural to think that the stroke judges are more likely to be fully certain of infractions in the end lanes (lanes 1 and 8 in an 8-lane pool) rather than in the middle lanes, which are further away from them. Since all swimmers are supposed to be judged equally, it seems like this setup may create an element of bias.

However, proving it from a mathematical or statistical standpoint is much more challenging. First, in high level competitions swimmers are much more experienced and stroke disqualifications are rare, making it difficult to observe and prove any bias towards end lanes. On the other hand, in lower-level competitions, both the judges and swimmers are often much less experienced, and standards vary significantly between competitions, making it much harder to draw general conclusions.

Data Collection and First Observations

This year's County championships offered a rare opportunity to study and quantify this phenomenon. It was a "level 1 competition", which means that the judges must be highly qualified and experienced. However, at this event some of the swimmers were not very experienced, making stroke infractions more frequent. The competition results were publicly posted on the county website, so I have gone through the data for the races held in an 8-lane pool. In total, there have been 38 disqualifications. Of those, I determined that 15 were submitted by stroke judges. Below is the summary data for those 15 disqualifications:

Session	Event #	Event Name	Gender	Swimmer first r	Age	Heat	Lane	End lan	Disqualification Reason
5	51	400 IM	Girls	Emily	15	4	3	No	7.2- stroke cycle not one arm stroke to one leg kick
5	51	400 IM	Girls	Alexia	16	4	1	Yes	8.3- movements of legs not simultaneous
5	55	200 Fly	Girls	Olivia	11	1	2	No	8.5- head not on surface during stroke
5	55	200 Fly	Girls	Ines	12	1	1	Yes	8.5- head not on surface during stroke
5	55	200 Fly	Girls	Ruby	12	1	7	No	8.5- head not on surface during stroke
5	55	200 Fly	Girls	Olivia	14	3	8	Yes	8.3- movements of legs not simultaneous
6	63	100 Fly	Boys	Finlay	11	2	8	Yes	8.5- head not on surface during stroke
6	63	100 Fly	Boys	Martin	12	Final	8	Yes	8.2- Non-simultaneous arms/ arms not above wat
6	63	100 Fly	Boys	Caspar	12	3	1	Yes	8.3- movements of legs not simultaneous
7	71	400 IM	Boys	Seth	14	6	8	Yes	7.1- Head not breaking surface when supposed to
7	71	400 IM	Boys	Konstantin	14	4	1	Yes	8.3- movements of legs not simultaneous
7	71	400 IM	Boys	Alfie	17	6	1	Yes	8.3- movements of legs not simultaneous
7	74	100 Breast	Girls	Alice	11	2	3	No	7.2- stroke cycle not one arm stroke to one leg kick
7	74	100 Breast	Girls	Maya	13	9	8	Yes	7.5- Alternating kick/ fly kick
7	75	200 Fly	Boys	Josh	13	1	6	No	8.3- movements of legs not simultaneous

The stroke disqualifications seem to be well distributed across different sessions, events and swimmers. The stroke judges rotated between sessions, so it is unlikely that one specific judge was in a particularly bad mood that weekend. Even though officiating at this competition was very strict, I am sure that all judges were fair and impartial. Swimming judges are volunteers, and usually they are parents of fellow swimmers. They would not disqualify a swimmer unless an infraction was clear and visible to them beyond doubt. I do not see any reason why they would intentionally give any preferential treatment to middle lanes over end lanes.

Quantitative Analysis of the Data

We can see that out of the 15 stroke disqualifications, a total of 10 were in the end lanes of the pool. To draw statistical conclusions from this observation, we will model this data using *Bernoulli trials*, a 300-year-old concept. In Probability and Statistics, a Bernoulli trial is a random experiment with two possible outcomes: *"success"* and *"failure"*. The probability of success remains the same each time the experiment is conducted, and the trials are independent. Jacob Bernoulli has analyzed it in his book about Combinatorics called *Ars Conjectandi*, which was published posthumously in 1713. The simplest example of a Bernoulli trial is flipping a coin, in which case probability of success, *p*, is 50%, unless the coin is biased.

If there is a stroke disqualification, we can think of "success" outcome as the affected swimmer having raced in the end lane of the pool. Considering the pool has 8 lanes, two of which are end lanes, if stroke judging is unbiased, the probability of "success" should be 25%. Note that this is a *conditional probability* and is assuming a disqualification has already occurred.

In our data, we had 15 trials and 10 "success" outcomes. So, what is the probability of such an event? To answer that question, we need to look at the binomial probability distribution, which gives the probability of *k* success outcomes in *n* independent trials, where the probability of success in each trial is *p*. The probability of this is given by:

$\mathrm{P}(\mathrm{X}=k)=\binom{n}{k}p^k(1-p)^{n-k}$, where $\binom{n}{k}=\frac{n!}{k!(n-k)!}$, and is also referred to as "*n* choose *k*", as well as sometimes being known as the *binomial coefficient*.

Intuitively, one can think of this formula as *k* successful trials having the probability of p^k, and the rest of the trials having to be failures, so for *n-k* failures the probability is $(1-p)^{n-k}$. Because we need both *k* successes *and n-k* failures to occur, their probabilities are multiplied. However, this represents only one combination of successes and failures out of many sequences. Since we do not care about the order of disqualifications, we need to include all possible combinations within *n* experiments. It turns out that the number of possible combinations is the binomial coefficient, a fundamental concept in combinatorics.

The binomial coefficient is the number of groups of k elements that can be chosen from a group of n elements, where the order in which the elements are selected does not matter (i.e. 12345 and 25143 are considered the same combination).

In our data of 15 trials and 10 disqualifications in the end lanes, the formula tells us that there are 15! / (10! x 5!) = 3003 ways to place 10 end lane disqualifications in 15 total without regard to order. The probability of having exactly 10 end lane disqualifications out of 15 total is then equal to

$\mathrm{P}(\mathrm{X}=10)=3003 * 0.25^{10} * 0.75^{5}=\mathbf{0.068}\%$**,** which is clearly very low.

We can also look at probabilities of all possible outcomes:

# of stroke DQs in end lanes	Probability
0	1.336%
1	6.682%
2	15.591%
3	22.520%
4	22.520%
5	16.515%
6	9.175%
7	3.932%
8	1.311%
9	0.340%
10	**0.068%**
11	0.010%
12	0.001%
13	0.000%
14	0.000%
15	0.000%

Binomial Hypothesis Testing

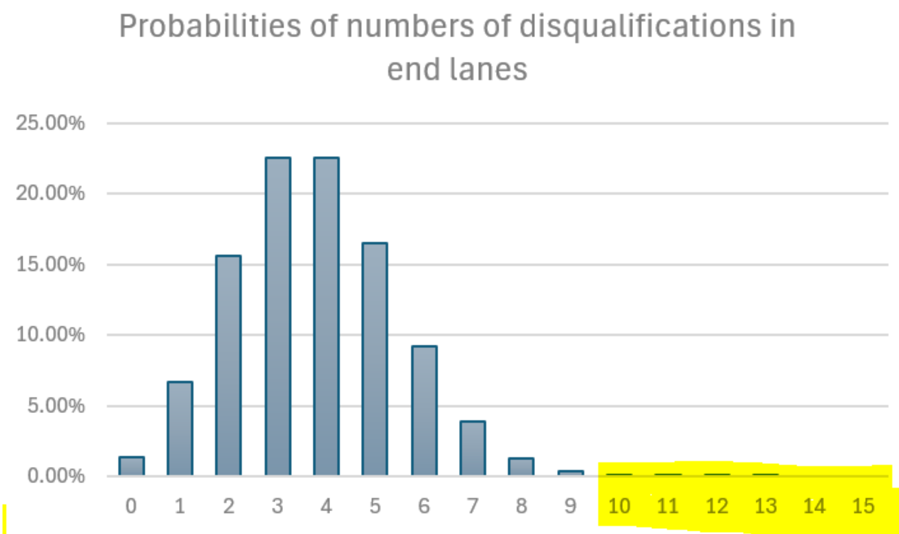
Does the specific outcome - 10 disqualifications out of 15 being in end lanes- tell us anything about potential bias? Or was this just "bad luck" for the end lane swimmers?

Another way of thinking about this is if someone flips a coin 10 times and all 10 times we see tails, we suspect that the coin is biased, but the question is how we know for sure. The answer is we do not, but we can know with some specific level of certainty. Quantifying that level of certainty is called *"Binomial Hypothesis testing"*.

If the disqualifications are completely independent of the lane number in an 8-lane pool, the probability *p* of a specific disqualification to apply to an end lane is 25%. This is called the "*Null hypothesis*". We can accept or reject this hypothesis at a "*significance level",* which is the probability that the observed result occurs by chance. For example, if significance level is 1%, then we are testing to see whether the probability of the *observed outcome or a more extreme outcome* is less than 1%. If that is the case, then the result is considered statistically significant, and the Null hypothesis can be *rejected* with 99% certainty.

In our case, the probability that there are 10 or more disqualifications in end lanes out of 15 total would be the sum of probabilities in the shaded area of the above table or the below graph, which is 0.08%. Therefore, with our assumptions, *we can safely reject the Null hypothesis that stroke disqualifications are independent of the lane the swimmer races in*. In fact, we could say that with 99.9% likelihood there is a bias towards end lanes in stroke disqualifications, at least in the case of the specific competition we have been analyzing and within the assumptions of our model.

Since we are only interested in one side of the distribution – success outcomes greater than 10, what we have performed is called a one-tailed binomial hypothesis test.



Probability distribution of end lane disqualifications, highlighting the tail we are testing.

We can also calculate the critical value, which is the lowest number of end lane disqualifications to result in rejection of Null hypothesis with 99% likelihood (if our significance level is 1%). We can see by looking at the data table that 9 or more disqualifications would produce the same result. However, if we were satisfied with 94% confidence, 7 or more disqualifications would result in us rejecting the null hypothesis.

Could we have reached the same conclusion by analyzing other competitions? Not as easily. There are usually very few, if any, stroke disqualifications in level 1 competitions. For example, at the London Region Summer Championships 2023 there has been a total of 2 stroke disqualifications, one of which came from an end lane. This does not tell us anything. Building out the binomial distribution for this competition, we will see that 1 end lane disqualification is not sufficient evidence to reject the Null hypothesis. We would only be able to do it with 56% certainty, which is clearly insufficient. That is why it is so difficult to analyze the data on this subject – in high level competitions there are simply too few stroke disqualifications to draw a credible conclusion, whereas at lower level the data is unreliable for reasons explained earlier. That is why the County Championships were so unique in that sense – highest level of officiating and many stroke disqualifications at the same time. In the table below we show how many disqualifications and success outcomes would be sufficient evidence for our analysis. I have calculated the probability that the stroke disqualifications were fully independent of the lane for different scenarios and "success" outcomes:

	Total number of stroke disqualifications														
# of end lane DQs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
1	25%	44%	58%	68%	76%	82%	87%	90%	92%	94%	96%	97%	98%	98%	99%
2		6%	16%	26%	37%	47%	56%	63%	70%	76%	80%	84%	87%	90%	92%
3			2%	5%	10%	17%	24%	32%	40%	47%	54%	61%	67%	72%	76%
4				0%	2%	4%	7%	11%	17%	22%	29%	35%	42%	48%	54%
5					0%	0%	1%	3%	5%	8%	11%	16%	21%	26%	31%
6						0%	0%	0%	1%	2%	3%	5%	8%	11%	15%
7							0%	0%	0%	0%	1%	1%	2%	4%	6%
8								0%	0%	0%	0%	0%	1%	1%	2%
9									0%	0%	0%	0%	0%	0%	0%
10										0%	0%	0%	0%	0%	0%
11											0%	0%	0%	0%	0%
12												0%	0%	0%	0%
13													0%	0%	0%
14														0%	0%
15															0%

We can see that for a low number of total disqualifications, it is impossible to draw any credible conclusions, and that in the area highlighted in red, the conclusion is obvious. This includes our case (10 end lane disqualifications out of 15 total). The same conclusion would apply even if we had 1 or 2 less end lane disqualifications in our data.

Adding the data from different competitions with only 1 or 2 stroke disqualifications in each would also be problematic, as the resulting set is likely to have many individual outliers, which are not relevant to our analysis. For example, when a swimmer does not finish the race (chokes in the middle of the pool, gets exhausted, etc) and gets disqualified, it is sometimes reported by a stroke judge as a stroke infraction.

Limitations and Challenges of the Model

I've tried to think of what could be wrong with our assumptions and methods. First, one may argue that the trials are not completely independent. There is usually a specific algorithm for placing swimmers in lanes, the most popular one is by entry time. Assume 100 swimmers are about to race in a specific event in an 8-lane pool. First, swimmers are sorted by their *entry time* – the fastest time they have achieved in this event in previous competitions. They are then placed into individual races called "heats". The last heat is made up of swimmers with 8 fastest entry times, second last are the next 8 fastest swimmers the process is repeated until there are less than 8 swimmers left, in our example 100 mod 8 = 4, and these swimmers are in heat 1 – slowest heat. There is an exception when there are only 1 or 2 swimmers left, in that case 3 swimmers are placed in heat 1 and 6 or 7 in heat 2. Within each heat, the swimmers are placed in lanes in the following order, from fastest to slowest entry times: 4, 5, 3, 6, 2, 7, 1, 8.

	Lane							
Heat	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
13	7	5	3	1	2	4	6	8
12	15	13	11	9	10	12	14	16
11	23	21	19	17	18	20	22	24
10	31	29	27	25	26	28	30	32
9	39	37	35	33	34	36	38	40
8	47	45	43	41	42	44	46	48
7	55	53	51	49	50	52	54	56
6	63	61	59	57	58	60	62	64
5	71	69	67	65	66	68	70	72
4	79	77	75	73	74	76	78	80
3	87	85	83	81	82	84	86	88
2	95	93	91	89	90	92	94	96
1			99	97	98	100		

Example of a placement of 100 swimmers into 13 heats. Swimmers are sorted by entry time from 1 (fastest) to 100 (slowest) with the first (slowest) heat swam with empty end lanes.

The question is whether this placement algorithm challenges our assumption of independence between the trials.

I would argue that it does not affect our analysis for the following reasons: (a) Based on the above method, the slowest heat (heat 1) is in 75% of the cases swum without end lanes, yet it consists of the slowest swimmers in the whole sample. Therefore, if there was any bias introduced by the assumption that slower swimmers are more likely to get disqualified, heat 1 would produce a lot more infractions in the middle lanes than in the end lanes. (b) With the above placement algorithm, lane 4 of heat *H* would have a weaker swimmer than lane 8 of heat *H+1*. Therefore, if stroke disqualifications were unbiased, we would argue that a middle lane swimmer in each heat would be more likely to be disqualified than lane 8 swimmer in a subsequent heat. That said, in certain competitions the heats are "circle-seeded", meaning the top ranked swimmers always get the center lanes in the last 3 heats. However, in those cases typically the level of swimmers in the last heats is high enough to have a very low probability of a stroke infraction.

The second potential challenge is that in our data, some of the stroke disqualifications may have in fact been done by a turn judge rather than a stroke judge, because if a stroke infraction was made within a few meters of the turn, the turn judge is the one to report it. During the data collection, I've done my best to remove all turn and other non-stroke infractions, but it is possible that one or two stroke infractions could be reported by a turn judge rather than a stroke judge. However, if we have erroneously included such disqualifications, it would in fact skew our data more towards the null hypothesis and support our conclusion even more, because each turn judge is positioned to observe a specific lane, as opposed to a stroke judge observing all the lanes.

Conclusion

Despite our approach having some challenges and limitations, we have statistically proven that there is a significant bias towards end lanes in stroke disqualifications. A natural question to ask is “why?”. We have already stated that the judges themselves are unbiased and do not intentionally favor any specific lanes. This pattern is likely observed because of their position on poolside being much closer to the end lanes and therefore being able to see them better. The rules state that the swimmer always gets the benefit of the doubt, so officials would not be able to disqualify middle lanes if they could not clearly see what happened. For example, minor alternating kick on butterfly or breaststroke is hard to see from far away, so middle lanes are much more likely to get away with this, even if stroke judges are giving equal attention to all lanes. Also, stroke officials are meant to walk closest to the slowest swimmers, who are most likely to be in end lanes due to how the heats are seeded.

Now, the final question is how we can eliminate this bias. First one is a video review, which is like VAR in football. It is currently used in very high-level competitions, such as World Championships, Olympics and NCAA championships. However, it is probably too expensive and impractical for lower-level competitions, and it often results in even harsher judgement. Another way to eliminate bias would be to change positioning of the officials. However, this is hard unless we can have a drone flying over the pool or place officials on a raised platform and equip them with binoculars. Finally, we could try to change the officiating standard and either make it harsher towards middle lanes, which would mean no benefit of the doubt to the swimmer in certain cases, or be more lenient towards end lanes, where, for example, the infraction must be so obvious that the official could be certain to see it even at a distance.

As for now, whenever I race in the end lanes, I just know to be more careful with the stroke rules.