# eDNA Effort Theory

*Ben Buzzee*

*May 20, 2019*

## Determining eDNA effort requirements

Our goal is to determine how many eDNA samples must be taken in order to achieve a given probability of detection. Dunker (2016) estimated the probability of detection at 1, 10 and 40 meters given the presence of a pike ($P(D|Z_i)$). So we will assume the probability of detection is known and constant for the circular regions 0-1m, 1-10m, and 10-40m from the test location. We will also assume the probability of detection given a pike is over 40m from the test site is zero, since no information is available for that region.

Let D be the event a single pike is detected and $Z_i =$ the event that the pike is in circular region $i$. Then we have that $P(D) = P(D \cap Z_1) + P(D \cap Z_2) + P(D \cap Z_3)$ by the law of total probability.

Next, by applying the definition of an intersection to the right hand side if the equation, we have that $P(D) = P(D|Z_1)P(Z_1) + P(D|Z_2)P(Z_2) + P(D|Z_3)P(Z_3)$

Finally, by assuming fish are randomly distributed across the entire lake, we can determine $P(Z_i) = \frac{area\ of\ region\ i}{total\ area}$

Thus for one sample, we can compute the probability of detection as $P(D_1) = P(D|Z_1)P(Z_1)+P(D|Z_2)P(Z_2)+P(D|Z_3)P(Z_3)$ where the conditional probabilities come from Dunker 2016 and the P($Z_i$)'s are the proportions of total area represented by the respective circular region.

Next, we need to expand this calculation to multiple samples. We can do this by assuming the tests and test sites are identical, or that for samples sites A and B, P(D|A) = P(D|B) = c.

Then it can be showed that $P(D|A \cup B) =$ c by applying bayes theorem, the law of total probability, and some fancy algebra. This holds in general for 1, 2,. . . , n sample sites. So all that changes in our above probability formula is that there are now S subregions, so the probability a pike is in one of S sub region i's is now $S * P(Z_i)$

That gives us the probability of detecting one pike at n separate sample sites as $P(D) = P(D|Z_1)(S)P(Z_1) + P(D|Z_2)(S)P(Z_2) + P(D|Z_3)(S)P(Z_3)$ If pike are independent, the number of detections given there are N pike in the system is a binom(N, P(D)) random variable.

The probability of at least one detection is $1 - P(no\ detections) = 1 - (1 - P(D))^N$