Akshaj Kammari
ak1990

Problem 1: The Future of Cream Cheese

1. From the information given about the poll, P(poll) = P(heads) + P(tails). P(heads) = P(tails) = 0.5 due to uniform distribution. The information about the poll goes even more in depth about P(tails), as it can be broken down into P(tails) = P(heads) + P(tails), which is also uniform. So essentially, P(poll) = P(heads) + (P(heads) + P(tails)), which numerically is: P(poll) = P(heads) + ¼(yes) + ¼(no). P(heads) can be broken down further as well, as it means the employee answers honestly so P(heads) = P(yes) + P(no). q is taken as the probability that the employee actually likes the strawberry cream cheese so P(yes) = q and P(no) = 1-q. p is the probability that the employee says that they like the strawberry cream cheese, which is inclusive of q. Hence, p = (P(heads) * P(yes)) + (P(tails) * P(heads)). Thus, p = (½ * q) + (½ * ½). p = ½q + ¼.

2. This question is specifically asking for the representation of p̂, in the formula of $\hat{p}_N$ = X/N, where X is the number of people who like strawberry cream cheese and N is the number of people polled. Using x to represent the number of specific outcomes, the distribution of $N * \hat{p}_N = p^x * (1 - p)^{N - x}$.

3. Computing $E[\hat{p}_N] = p$ shows that p̂ is an unbiased estimator. Substituting $\hat{p}_N = X/N$ gives us E[X/N], which can be simplified even further, knowing E[X] = N * p, simply making E[X/N] = N * p / N = p, making it unbiased. $Var(\hat{p}_N) = E[\hat{p}_N^2] - E[\hat{p}_N]^2$. Since we know $e[\hat{p}_N] = X/N$, making $E[\hat{p}_N^2] = E[(X/N)^2] = E[X^2] / N^2$. The $E[X^2]$ part can be rewritten as E[X]*E[X - 1] + E[X]. E[X] = N*p, which can be substituted in the most recent expression, resulting in (N*p)*((N-1))*p) +N*p. Not forgetting about the N^2, $E[\hat{p}_N^2] = ((N - 1) p^2 + p) / N$. $E[\hat{p}_N] = p$, so $E[\hat{p}_N]^2 = p^2$. Plugging it in, we get $Var(\hat{p}_N) = (((N - 1) p^2 + p) / N) - p^2 = p * (1 - p) / N$.

4. We previously got that p = ½q + ¼. Solving for q: q = 2p - ½. A random variable may be set to $\hat{q}_N = 2\hat{p}_N^2 - ½$. Solving for $E[\hat{q}_N]$, $E[\hat{q}_N] = E[2\hat{p}_N - ½] = 2E[\hat{p}_N] - E[½]$. We had also declared that $E[2\hat{p}_N] = p$, leaving us with $E[\hat{q}_N] = 2p - ½ = q$.

5. We can use the equation $Var(\hat{q}_N) = E[\hat{q}_N^2] - E[\hat{q}_N]^2$. We can further simplify this by subbing in values of $\hat{q}_N = 2\hat{p}_N^2 - ½$ and E[q̂N] = 2p - ½ = q into $E[\hat{q}_N^2]$ and $E[\hat{q}_N]^2$. $Var(\hat{q}_N) = (4p^2 - 2p + 1/3) - (2p - 1/2)^2 = (4p^2 - 2p + 1/3) - (4p^2 - 2p + 1/4) = 1/12$.

5. To find the variance of Var($\hat{q}_N$), we can simply substitute the value of p = ½q + ¼ into the equation Var(p̂N) = p * (1 - p) / N, resulting in an expression in terms of q and N as follows: Var($\hat{q}_N$) = (½q + ¼)^2 * (1 - (½q + ¼)) / N = (2q + 1)^2 * (-2q + 3) / 64N.

6. To find out how many people to poll, the minimum sample size must be calculated taking into consideration the level of confidence, which is 95%, and the margin of error, which is 0.01. The margin of error E may be able to be expressed as E = x * sqrt[($\hat{q}_N$(1 - $\hat{q}_N$)) / N]. x corresponds to the desired level of confidence, so for 95% confidence, x = 1.96. Since E should not be more than 0.01 1.96 * sqrt[($\hat{q}_N$(1 - $\hat{q}_N$)) / N] <= 0.01. N >= (1.96^2 * $\hat{q}_N$ * (1 - $\hat{q}_N$)) / (0.01^2). Using a conservative estimate due to the fact of no known true value of q, we assume half of the employees want strawberry cream cheese. Plugging that into the equation, N >= (1.96^2 * 0.5 * 0.5) / (0.01^2) ≈ 9604: we need to poll at least 9604 people using the coin-flipping method to achieve a 95% confidence level and a margin of error of 0.01.

Bonus: We may also calculate the sample size required for natural polling using this formula N = (x^2 * p * (1 - p)) / E^2. Plugging in values also results in N = (1.96^2 * 0.5 * 0.5) / (0.01^2) ≈ 9604, which means that the sample size is the same.

Problem 2: Strawberrymandering

1. The question asks for the probability of SCCI passing with a group of different amounts of people. 17 of the 33 voters would have to support it for it to pass, hence, binomial distribution may be used to calculate the probability of P(T >= 17), which is ∑N choose x * p^x * (1 - p)^T - x. x represents 17-33 and N is the total number of votes cast. This results in a probability of 0.2808 for 33 voters. 152 of the 303 votes need to support SCCI for it to pass and using the same logic as before, it results in the probability of 0.0513. Finally, 1502 of the 3003 votes need to support RCCI to pass, but this probability is so small it can be considered negligible.

2. The total probability of RCCI passing is (p1 + p2 + p3) / 3 = 0.45, which can be further simplified as p1 + p2 + p3 = 1.35.

3. The probability of yes would still be 0.45 even after being split. If we take n to be N/3 though, the probability may be higher because of the change in number of trials using the same formula as before. For each P(i), we can calculate the probability of a district P(i) to pass by using the formula ∑(n choose k) p^k(1 - p)^n - k. k represents the majority and sum up til N/3. The probabilities of vote passing can be calculated as the outcomes for at least 2 districts passing SCCI, which is p1 & p2, p2 & p3, p1 & p3, or all of them.

4. Using the formula $\sum$(n choose k) $p^k(1 - p)^n$ - k, where k is the number right above majority and n is the total votes, the optimal choice is where the probability remains the highest. For the district having 11 votes, 6 is needed for majority resulting in 0.367 for the optimal probability for each district to pass. Higher votes than 6 means a lower probability and lower than 6 means the district will not pass. This formula may be reused, and because there are 3 choose 2 possible outcomes for SCCI to pass, using p = 0.367, $\sum$(3 choose 2 for 11 votes) $p^k(1 - p)^n$ - k = 0.305. Repeating this for 101 votes results in 0.152 and for 1001 votes is negligible as it is ~ 0.

Bonus: The optimal level of support may equal the derivative for each district set equal to 0. The number of votes cast in each district may be an important factor as it could help find the optimal number of votes supporting.

Problem 3: Splitting the Atom

1. The problem is asking for if the atom is undecayed at time >= 1, or in other words P(Xi(t) = 1). (1 - p) represents the probability that atom i has not decayed yet, and the probability of the decay of the atom at t time would have us raise (1-p) to the time t, resulting in $(1 - p)^t$.

2. We can use the same formula from the previous question, $(1 - p)^t$, representing the probability that the atom stays undecayed. To solve this problem, we can simply multiply it by the number of initial atoms N, resulting in $N *(1-p)^t$.

3. To find $t_{1/2}$ we need to declare a decay constant, c, due to the fact that this will not be affected by the original amount of material. This will depend on the isotope. The equation is $t_{1/2} = \ln(2)/(1 - p)$.

Bonus: Due to the analysis given, the upper limit is about 50,000 years regarding accuracy of the carbon dating. This limit is due to the fact that Carbon dating makes use of half-lives, which is essentially the time it takes for the carbon present to half itself due to decay. As time passes, the Carbon-14 present decays more and more until there is too less of it left to measure, diminishing the accuracy of the measurement.

Problem 4: Where is Max?

1. The probability of finding Max on the right floor must be combined with finding Max on floor i as well as randomly selecting that floor, which can be converted into a summation Σ(1 to n) $P(p_k) * P(q_k)$.

2. qj should be proportional to P(i) in order to maximize the probability of saving Max, using this formula: $(1/ \Sigma P(i)) * P(i)$.

3. For the highest probability of success for us, all p_i must be equal resulting in the formula $\Sigma(1 \text{ to } n)\ 1 / N$.

4. This equation should be used to maximize the probability for saving Max: $\Sigma(1 \text{ to } n)\ 1 / N$.