KANGHWI LEE

CS 206

BagelBot and the Theories of Mind, Games, and the Atom 16:198:206

**Problem 1: The Future of Cream Cheese (30 points)**

1) Let q be the probability a random employee actually likes strawberry cream cheese, and let p be the probability that employee responds to the poll saying that they like strawberry cream cheese. What’s the relationship between p and q?

**- N!**

2) Let X be the number of people who say yes to strawberry cream cheese via the poll, out of N people polled. (Note: If this ends up on Chegg, someone has let me down tremendously, and I hope everyone involved sits and thinks about the choices they made.) Let ˆpN = X/N, i.e., the fraction of people who polled in support of strawberry cream cheese. What is the distribution of N ∗ pˆN ?

**- P(X=k) = (N choose k) p^k (1-p)^(N-k)**

**- Binomial distribution with parameters N and pN**

3) Show that ˆpN is an unbiased estimator for p. What is E [ˆpN ]? What is Var(ˆpN )?

**- E(pN) = p**

**Var(pN) = Var(X/N) = 1/N^2 \* Var(X) = 1/N^2\*(1-p)NP = (1-p)p/N = Var(^pn)**

**- PN is an unbiased estimator for p. E[PN] = p, Var(PN) = p(1-p)/N**

4) Construct a random variable ˆqN based on the polling data (N, X, pˆN ), such that E [ˆqN ] = q.

**- The expected value of X is (1/2)qN + (1/2)(1/2)N = (1/4)N + (1/2)qN, and the estimated value of q is (2X-N)/N.**

**- qHat = 2(pHat) – 1/2**

5) What is the variance of ˆqN , in terms of N and q?

**- Var(^qN) = Nq(1-q)/N^2 = q(1-q)/N**

6) How many people should you poll if you want 95% confident your estimate ˆqN is within 0.01 of the true value of q? Note: since you are trying to find q, the number of people you are going to poll can’t be based on q!

**- Margin of error (E) = zsqrt(p(1-p)/n)**

**- 1.96sqrt(0.5(1-0.5)/n) <= 0.01**

**Solving for n:**

**n >= (1.96^2 \* 0.5 \* (1-0.5)) / 0.01^2**

**n >= 9604**

Bonus) How many more people do you have to poll using this method to get accurate results (within ±0.01 of the true value of q with 95% confidence), compared to just polling naturally and analyzing the results?

* **X >= 9604**

**Problem 2: Strawberrymandering (20 points)**

1) If a poll were held right now to add or not add Strawberry Cream Cheese, based on majority vote, what’s the probability SCCI passes, based on 33 total votes? 303 total votes? 3003 total votes?

**- P(SSCI Passes)= C (n, k) ⋅ pk ⋅ (1 - p) (n - k)**

where n is the number of votes cast, k is the number of votes cast in favor of the bill, and p is the likelihood that a single senator will vote in favor of the bill.  
For n=33:

**P(SSCI Passes)= C(33, k) ⋅ pk ⋅ (1 - p) 33 (- k)**

* **0.479**

**For n = 303:**

**P(SSCI Passes)= C(303, k) ⋅ p ⋅ (1 - p) 303 (-k)**

* **0.508**

**For n = 3003:**

**P(SSCI Passes)= C(3003, k) ⋅ pk ⋅ (1 - p) 3,003 (-k)**

**- 0.518**

2) Let p1, p2, p3 be the probability of support for the SSCI in Districts 1, 2, 3 respectively. (Note: If this ends up on Chegg, someone has let me down tremendously, and I hope everyone involved sits and thinks about the choices they made.) If each district has the same total population, and the level of support across the entire group of employees is 0.45, what has to be true about p1, p2, p3?

**- (p1+p2+p3)/3 = 0.45**

3) In terms of p1, p2, p3, if there are n votes cast in each district, what’s the probability that the SSCI passes?

**- P(at least 2 districts pass) = 1 - P(no districts pass) - P(1 district passes)**

**where**

**P(no districts pass) = (1-p1)^n \* (1-p2)^n \* (1-p3)^n**

**P(1 district passes) = n \* [p1\*(1-p2)(1-p3) + (1-p1)p2(1-p3) + (1-p1)(1-p2)\*p3]**

4) What is the optimal choice for p1, p2, p3? For these districts, what’s the probability that the SSCI passes with 11 total votes in each district? 101 votes in each district? 1001 votes in each district?

**- p1 p2 p3 should be 0.45 for all of them.**

Bonus) What would the optimal levels of support be if there were four districts? Five districts? k districts?

* **Let p be the proportion of voters who support the candidate in each district, and let k be the number of districts.**

**If the candidate wins a majority of the districts, they will win the election. Therefore, the optimal levels of support would be those that maximize the probability of winning a majority of the districts. The probability of winning i out of k districts is given by the binomial distribution:**

**P(i) = (k choose i) \* p^i \* (1-p)^(k-i)**

**where (k choose i) is the binomial coefficient.**

**To maximize the probability of winning a majority of the districts, we need to find the value of p that maximizes the probability of winning at least half of the districts. That is, we need to find the value of p that maximizes the cumulative distribution function (CDF) of the binomial distribution:**

**CDF(i) = sum(P(j) for j=i to k)**

**where i is the number of districts won and k is the total number of districts.**

**Finding the optimal value of p analytically can be difficult, but we can use numerical optimization methods to approximate it.**

**Problem 3: Splitting the Atom (30 Points)**

1. What is the probability that atom i is undecayed at time t ≥ 1, i.e., P(Xi(t) = 1)?

**- The probability that a specific atom will remain undecayed at a given time.**

**- P(Xi(t)=1)=(1−p)^t**

1. Find µ(t) = E [X(t)], the expected number of strontium-90 atoms that are left at time t. (Note: If this ends up on Chegg, someone has let me down tremendously, and I hope everyone involved sits and thinks about the choices they made.) Show that this decays exponentially, at a rate determined by p.

**- The expected number of strontium-90 atoms that will remain at a given time, which decays exponentially with a rate determined by the probability of decay.**

**- u(t)=N(1−p)^t, which is an exponential decay with rate p.**

3) At what time t is the expected amount of strontium-90 left no more than half the original amount? Call this the half life, t1/2, and show it doesn’t depend on the original amount of material. The previous calculations are meant to show that the expected or average amount of radioactive material left at a given time behaves in a very predictable manner with an exponential decay. However, what about the actual amount of radioactive material?

**- The half-life of strontium-90, which is the time it takes for the expected amount of strontium-90 to decay to half of its original amount, and it doesn't depend on the original amount of material.**

**- t1/2=ln(2)/p, which is independent of N.**

4) Derive an upper bound on P(X(t) ≥ µ(t)(1 + )) in terms of N, , p, and t. This is the probability that the actual amount of material left at time t is larger than the expected amount by a factor of 1 + .

**- An upper bound on the probability that the actual amount of material left at a given time is larger than the expected amount by a factor of 1 + e.**

**- P(X(t)>u(t)(1+e)) ≤ e(−2e^2Np(t+1))**

5) Derive an upper bound on P(X(t) ≤ µ(t)(1 − )) in terms of N, , p, and t. This is the probability that the actual amount of material left at time t is smaller than the expected amount by a factor of 1 − .

**-  An upper bound on the probability that the actual amount of material left at a given time is smaller than the expected amount by a factor of 1-e.**

**- P(X(t)≤u(t)(1−e)) ≤ e(−2e^2Np(t+1))**

6) Noting the number of atoms in a scoop of strontium-90 is about N ≈ 1023, show that with probability almost 1, X(t1/2) is between 0.999µ(t1/2) and 1.001µ(t1/2). As such, the apparent contradiction seems resolved - while it is the expected amount of radioactive material that decays exponentially in a very predictable fashion with a known half life, the actual amount of radioactive material (though random) is concentrated around this with very high probability.

**- It shows that with almost 100% probability, the number of undecayed atoms in a scoop of strontium-90 at half of its half-life is very close to the expected number of undecayed atoms, which was calculated in part 3.**

**- P(0.999μ(t1/2)≤X(1/2 ) ≤ 1.001μ(t1/2)) ≥ 1−e(−2.5×10^1 5)(withN≈10^2 3)**

Bonus) All living organisms that we know of contain Carbon. A fraction of this carbon is Carbon-14, a radioactive isotope of Carbon. The proportion of Carbon-14 in an organism is constant over its lifespan, as the organism is constantly replenishing the carbon in its body. However, after the organism dies, the Carbon-14 decays at a known rate, allowing scientists to gauge how long ago the organism was alive based on the amount of Carbon-14 remaining. However, this is not accurate beyond a timescale of about 50, 000 years. Based on the above analysis, why might there be an upper limit on how long a timescale this technique could be applied over?

* **The bonus part of the problem relates to the use of Carbon-14 to determine the age of organic materials**. **The problem explains why there is an upper limit on how long this technique can be used to determine the age of materials.**
* **It explains why the carbon-14 dating technique has a limit of about 50,000 years. This is because the amount of carbon-14 left after 50,000 years is so small that it's difficult to measure accurately, which means that the probability of error in the measurement is very high.**
* **Finally, the problem shows that, with a very high probability, the number of undecayed atoms at half the half-life time will be between 0.999 and 1.001 times the expected value. (μ(t1/2) and 1.001μ(1/2))**

**Problem 4: Where is Max? (20 Points)**

1. If the pj , qj probabilities are fixed in advance, what is the probability that you pick the right floor and save Max?

**- P( correct floor)=P(correct floor) = Σi (pi×qj)**

2) If you knew the probabilities pi (Note: If this ends up on Chegg, someone has let me down tremendously, and I hope everyone involved sits and thinks about the choices they made.) , what probabilities qj should you use to maximize your probability of saving Max? But you know that MagusCorp is clever. You have to assume they know you’re coming to rescue him, and will take that into account.

**- To use the Bayes' rule:**

**qj = / Σi pj**

3) How should MagusCorp choose the probabilities pi to minimize your best possible probability of saving Max?

**- pi = qj / (q1+q2+...+qN) if j is the floor with the highest probability qi among all floors i=1,2,...,N; otherwise pi = 0.**

4) Assuming MagusCorp hides Max according to the above probabilities, what is the probability you are going to be able to save Max

**- P(success) =1/N × Σi(pi×qj)**

**Substituting pi = 1/Nand qj= pi/Σi×pi, we get:**

**P(success) = 1/N × Σi × 1N/Σi×(1/N)**

**P(success) = 1/N**