

Problems for Recitation 22

Suppose that a coin that comes up heads with probability p is flipped n times. Then for all $\alpha < p$

$$\Pr \{\# \text{ heads} \leq \alpha n\} \leq \frac{1 - \alpha}{1 - \alpha/p} \cdot \frac{2^{nH(\alpha)}}{\sqrt{2\pi\alpha(1-\alpha)n}} \cdot p^{\alpha n}(1-p)^{(1-\alpha)n}$$

where:

$$H(\alpha) = \alpha \log_2 \frac{1}{\alpha} + (1 - \alpha) \log_2 \frac{1}{1 - \alpha}$$

1 Approximating the Cumulative Binomial Distribution Function

A coin that comes up heads with probability p is flipped n times. Find an upper bound on

$$\Pr \{\# \text{ heads} \geq \beta n\}$$

where $\beta > p$. Think about the number of tails and plug into the monster formula above.

A Gallup poll found that 45% of the adult population of the United States plan to vote Republican in the next election. Gallup polled 640 people and claims a margin of error of 3 percentage points.

Note that the only randomization in this experiment is in who Gallup chooses to poll. So the sample space is all sequences of n adult Americans. The response of the i -th person polled is “yes” with probability p and “no” with probability $1 - p$ since the person is selected uniformly at random. Furthermore, the n responses are mutually independent.

- a. Give an upper bound on the probability that the poll's estimate will be 0.04 or more too low. Just write the expression; don't evaluate yet!

- b. Give an upper bound on the probability that the poll's estimate will be 0.04 or more too high. Again, just write the expression.

- c. The sum of these two answers is the probability that Gallup's poll will be off by 4 percentage points or more, one way or the other. Unfortunately, these expressions both depend on p — the unknown fraction of voters planning to vote Republican that Gallup is trying to estimate!

However, the sum of these two expressions is maximized when $p = 0.5$. So evaluate the sum with $p = 0.5$ and $n = 640$ to upper bound the probability that Gallup's error is 0.04 or more. Pollsters usually try to ensure that there is a 95% chance that the actual percentage p lies within the poll's error range, which is $q \pm 0.04$ in this case. Is Gallup's poll properly designed?

3 Noisy Channel

Suppose we are transmitting packets of data across a noisy channel. Each packet has probability .01 of being lost. Now suppose we are transmitting 10,000 packets. What is the probability that at most 2% of the packets are lost?