Deviations from the Mean, Entropy Typical Set

Submit a PDF of your answers to Canvas

1. This problem we will explore bounding the probability that the sample mean deviates from the true mean for a Bernoulli random variables. Consider a sequence of independent coin flips, denoted $X_1, X_2, \ldots, X_n$. The coin flips are i.i.d. Bernoulli random variables:

$$X_i = \begin{cases} 0 & \text{with probability } 1-\theta \quad \text{tails} \\ 1 & \text{with probability } \theta \quad \text{head} \end{cases}$$

You flip the coin n times and observe k heads (where heads corresponds to $X_i = 1$).

a) Find/state the expression for $\widehat{\theta}_{\text{ML}}$, i.e., the maximum likelihood estimate of θ. $\hat{\theta} = k/n$

b) Find an exact expression for the probability that $\widehat{\theta}_{\text{ML}}$ deviates from the true mean by more than $\delta \in [0,1]$:

$$\mathbb{P}(|\theta - \widehat{\theta}_{\text{ML}}| \geq \delta)$$

Recall the Binomial pmf, which gives the probability of k heads in n coin flips:

Σ k lies within $(\theta-\delta)N, (\theta+\delta)N$

$$\binom{n}{k}\theta^k(1-\theta)^{n-k}$$

$1 - \sum_{k=(\theta-\delta)n}^{(\theta+\delta)N} \binom{n}{k}\theta^k(1-\theta)^{n-k}$

$\mu = \theta$, $\hat{\mu} = k/n$ $a, b = 0, 1$

c) Use Hoeffding's inequality to derive a bound on the probability that $\widehat{\theta}_{\text{ML}}$ deviates from the true mean by more than δ. $P(|\hat{\mu} - \mu| \geq \delta) = 2e^{\frac{-2\delta^2 n}{1}}$

d) Use a computer to evaluate the exact expression and Hoeffding's inequality when $\theta = 0.4$, $n = 10000$, and $\delta = 0.1$. both are astronomically small (exact $\sim \times 10^{-96}$ Hoeffding $\sim \times 10^{-87}$)

e) *Typically*, one would expect that the number of heads (after n flips) would be about θn. One precise notion of *typical* is based on the Shannon information. Recall that if the outcome of a sequence of coin flips is $x_1, \ldots, x_n$, the Shannon information (in bits) is given by $\log_2(1/p(x_1, \ldots, x_n))$. If X_i are independent, the Shannon information is $\sum_i \log_2(1/p(x_i))$. Define the set of outcomes that have approximately average Shannon information as the *entropy* typical set:

$$\left\{(x_1, \ldots, x_n) : \left|\frac{1}{n}\log_2\left(\frac{1}{p(x_1, \ldots, x_n)}\right) - H(X_i)\right| \leq \varepsilon\right\}.$$

Find an upper bound on the probability that an outcome is not in the entropy typical set by using Markov/Chebyshev's inequality.

Want $\left|\underbrace{\frac{1}{n}\log_2\left(\frac{1}{p(x_1, \ldots x_n)}\right)}_{\hat{\mu}} - \underbrace{H(x)}_{\mu}\right| > \varepsilon \leq \frac{\sigma^2}{n\varepsilon^2}$

$H(x) = E(-\log(x)) \rightarrow \sigma^2 = \frac{(-\log_2(0.4) + \log_2(0.6))^2}{12}$