Algorithm Foundations of Data Science and Engineering
Welcome Tutorial :-)
Tutorial 1

GAO Ming

DaSE @ ECNU

25 Feb., 2019

Tutorial 1

1. Let X be a r.v., $\mu = E(X)$ and $\sigma^2 = E[(X-\mu)^2]$. If $X^* = \frac{X-\mu}{\sigma}$, please prove $P[|X^*| \geq c] \leq \frac{1}{c^2}$.
2. Let X_i $(i = 1, 2, \cdots, n)$ be i.i.d., $\mu = E(X_i)$ and $\sigma^2 = E[(X_i - \mu)^2]$. If $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, please prove $P[|\overline{X} - \mu| \geq \varepsilon] \leq \frac{\sigma^2}{n\varepsilon^2}$.
3. In n tosses of a fair coin, let X be the number of heads, what's the probability of $X < \frac{n}{4}$ heads?
4. Let X_i be a sequence of independent r.v.s with $P(X_i = 1) = p_i$ and $P(X_i = 0) = 1 - p_i$. r.v. $X = \sum_{i=1}^{n} X_i$ and $\mu = \sum_{i=1}^{n} p_i$. Please prove the following conclusions.
   - $P(X > (1+\delta)\mu) < \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$
   - $P(X > (1+\delta)\mu) < \exp(-\mu\delta^2/4)$

Tutorial 1 Cont'd

5. For the situation of our running example (8 billion bits, 1 billion members of the set S), calculate the false-positive rate if we use three hash functions? What if we use four hash functions?
6. Suppose we have n bits of memory available, and our set S has m members. Instead of using k hash functions, we could divide the n bits into k arrays, and hash once to each array. As a function of n, m, and k, what is the probability of a false positive? How does it compare with using k hash functions into a single array?