CS70 HW 11

March 15, 2021

1 Random Cuckoo Hashing

(a) Pr(no collisions over the entire process) =

$$\frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{1}{n} = \frac{n!}{n^n}.$$

When n gets very large, it tends toward 0.

(b) For D_n, we have two cases: No collision with prob $\frac{1}{n}$. Collision with prob $1 - \frac{1}{n}$. Thus applying the law of total Expectation, we have

$$\begin{aligned} E(X) &= \frac{1}{n} \times E(0) + (1 - \frac{1}{n}) \times (1 + E(X)) \\ &= 1 - \frac{1}{n} + E(X) - \frac{1}{n}E(X) \\ \frac{1}{n}E(X) &= 1 - \frac{1}{n} \\ E(X) &= n - 1 \end{aligned}$$

2 Markov's Inequality and Chebyshev's Inequality

var(X) = 9, E(X) = 2 and Pr(x > 10) = 0. Thus we can transform this R.V to Y where first subtract X by 10 and flip (× -1). Then we have var(Y) = 9, E(Y) = 8 and Pr(x > 0) = 1.

(a) True.

$E(X^2) - E(X)^2 = E(X^2) - 4 = 9 \implies E(X^2) = 13.$

(b) True.

By Markov's Inequality, $P(X \leq 1) = P(Y \geq 9) \leq \frac{E(Y)}{9} = \frac{8}{9}$.

(c) Ture.

By Chebyshev's Inequality, $P(X \geq 6) = P(Y \leq 4) \leq P(\mid Y - 8 \mid \geq 4) \leq \frac{var(Y)}{4^2} = \frac{9}{16}$.

(d) False.

This R.V Y needn't to be symmetry, which may lead to $P(X \geq 6) > 9/32$. But I can't give a concrete counter-example.

3 Easy A's

(a) We define following events: S: total score; M: first Homework score; N: second Homework score; X: a distribution with mean $\mu = 5$ and variance $\sigma^2 = 1$. And we know N and M are independent and S = M + N, M = 3X, N = 4X.
Now we compute E(S) = E(M + N) = E(M) + E(N) = 3E(X) + 4E(X) = 7μ = 35. Var(S) = Var(M + N) = Var(M) + Var(N) = 9Var(X) + 16Var(X) = $25\sigma^2$ = 25.

(b) Still by Chebyshev's inequality, $Pr(S \geq 60) \leq Pr(|S - E(S)| \geq 25) \leq Var(S)/25^2 = 25/25^2 = 0.04 = 4\%$.

4 Confidence Interval Introduction

(a) $P(|X - \mu| \geq \varepsilon) \leq var(X)/\varepsilon^2 = \frac{\sigma^2}{\varepsilon^2}$.

(b) $|X - \mu| \geq \varepsilon \Leftrightarrow -\varepsilon < X - \mu < \varepsilon \Leftrightarrow X - \varepsilon < \mu < X + \varepsilon$. Therefore, $Pr(|X - \mu| \geq \varepsilon) = Pr(X - \varepsilon < \mu < X + \varepsilon)$.

(c) We compute as following:

$$
\begin{aligned}
P\{\mu \in (X - \varepsilon, X + \varepsilon)\} &= 1 - Pr(|X - \mu| \geq \varepsilon) \\
&\geq 1 - \frac{\sigma^2}{\varepsilon^2} \\
&\geq 0.95
\end{aligned}
$$

Then simplify the equation, we obtain

$$
\begin{aligned}
\frac{\sigma^2}{\varepsilon^2} &\leq 0.05 \\
\varepsilon &\geq \sqrt{20} \cdot \sigma.
\end{aligned}
$$

(d) $E(\bar{X}) = E(\frac{1}{n} \cdot \sum_{i=1}^{n} E(X_i)) = \frac{1}{n} \cdot n \cdot E(X_i) = \mu$.
$Var(\bar{X}) = Var(\frac{1}{n} \cdot \sum_{i=1}^{n} Var(X_i)) = (\frac{1}{n})^2 \cdot n \cdot Var(X_i) = \frac{1}{n} \cdot \sigma^2$.

(e) Same as in part(c), we want

$$
\begin{aligned}
\frac{Var(\bar{X})}{\varepsilon^2} &\leq 0.05 \\
\frac{\sigma^2}{n\varepsilon^2} &\leq 0.05 \\
\varepsilon^2 &\geq 20 \cdot \frac{\sigma^2}{n} \\
\varepsilon &\geq \sqrt{20} \cdot \frac{\sigma}{\sqrt{n}}
\end{aligned}
$$

So from the result, we can see $\varepsilon \to \frac{1}{\sqrt{n}}$, so as n gets large, ε gets small.