csci-5454-hw5

Jake Krol

October 2024

1 Question 1

1.1 Question 1a

Since jobs are assigned uniformly at random across all n servers, the probability of the jth job being assigned to server 1 is $\frac{1}{n}$, and the complementary case is $\frac{n-1}{n}$. These probabilities correspond to the two cases of the indicator random variable Y_j

$$Y_j = \begin{cases} 1 & \text{if job } j \text{ assigned to server 1} \\ 0 & \text{else} \end{cases}$$

Therefore, $Pr[Y_j = 1] = \frac{1}{n}$ and $Pr[Y_j = 0] = \frac{n-1}{n}$. The expected value of the discrete random variable Y_j is

$$\mathbf{E}[\mathbf{Y_j}] = \sum_{j=0}^{1} y_j Pr[y_j] = 0 \cdot \frac{n-1}{n} + 1 \cdot \frac{1}{n} = \frac{\mathbf{1}}{\mathbf{n}}$$

Y is a sum of indicator random variables and by linearity of expectation,

$$\mathbf{E}[\mathbf{Y}] = E[Y_1] + \ldots + E[Y_n] = n \cdot \frac{1}{n} = \mathbf{1} \tag{1}$$

1.2 Question 1b

$X = X_1 + \ldots X_n$ where X_i is an indicator variable and $Pr[X_i = 1] = \ldots = Pr[X_n = 1] = p$. Therefore, $E[X] = n \cdot p$, and in the context of part a $E[X] = n \cdot \frac{1}{n} = 1$.

Using Chernoff's multiplicative inequality $(Pr[X > (1 + \delta\mu)] \leq e^{-(\frac{\mu\delta}{2} \cdot ln(\delta)})$, bounding $Pr[\text{Server 1 gets more than } c \cdot \frac{ln(n)}{ln(ln(n))}] \leq \frac{1}{n^2}$ corresponds to setting $c \cdot \frac{ln(n)}{ln(ln(n))} = \delta$ since $\mu = 1$ and the additional 1 is negligible for large n.

$$
\begin{aligned}
(1+\delta)\mu &= c \cdot \frac{ln(n)}{ln(ln(n))} \\
(1+\delta)1 &= c \cdot \frac{ln(n)}{ln(ln(n))} \\
1+\delta &= c \cdot \frac{ln(n)}{ln(ln(n))} \\
\delta &= c \cdot \frac{ln(n)}{ln(ln(n))} - 1 \\
\delta &\approx \mathbf{c} \cdot \frac{\mathbf{ln(n)}}{\mathbf{ln(ln(n))}}
\end{aligned}
$$

Substituting in δ for the Chernoff bound,

$$
\begin{aligned}
Pr[X > c \cdot \frac{ln(n)}{ln(ln(n))}] &\leq e^{-1(\frac{\frac{ln(n)}{ln(ln(n))}}{2}) \cdot ln(\frac{ln(n)}{ln(ln(n))})} \\
&\leq e^{-\frac{c \cdot \frac{ln(n)}{ln(ln(n))}}{2} \cdot ln(c \cdot \frac{ln(n)}{ln(ln(n))})} \\
&\leq e^{-c \cdot \frac{ln(n)}{2ln(ln(n))} \cdot ln(c \cdot \frac{ln(n)}{ln(ln(n))})}
\end{aligned}
$$

The dominant term in the exponential for large n is $ln(n)$ compared to the $ln(ln(n))$ terms, so the exponential can be approximated to

$$
\begin{aligned}
Pr[X > c \cdot \frac{ln(n)}{ln(ln(n))}] &\leq e^{-\frac{c}{2} ln(n)} \\
&\leq e^{ln(n^{-\frac{c}{2}})} \\
&\leq \frac{1}{n^{\frac{c}{2}}}
\end{aligned}
$$

To ensure $\frac{1}{n^2}$, c must be 4.

$$
\begin{aligned}
\frac{1}{n^{\frac{x}{2}}} &= \frac{1}{n^2} \\
n^{\frac{-c}{2}} &= n^{-2} \\
\log_n(n^{\frac{c}{2}}) &= \log_n(n^{-2}) \\
-\frac{c}{2} &= -2 \\
\mathbf{c} &= \mathbf{4}
\end{aligned}
$$

2 Question 2

2.1 Question 2a

Y is a Bernoulli random variable with $p = \frac{1}{2}$ and $Pr[Y = 0] = Pr[Y = 1] = \frac{1}{2}$. The expected value of Y is

$$E[Y] = \sum_{j=0}^{1} p(y_j) \cdot y_j = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$$

Since Z depends on Y, X can be rewritten in terms of only Y by substituting $Y + 1$ for Z.

$$X = Z + Y = (Y + 1) + Y = 2Y + 1$$

By linearity, multiple, and constant rules of expectation,

$$\mathbf{E}[\mathbf{X}] = E[2Y + 1] = 2E[Y] + E[1] = 2 \cdot \frac{1}{2} + 1 = \mathbf{2}$$

Next, variance can be computed by $VAR[X] = E[X^2] - E[X]^2$ and since $X = 2Y + 1$ this means (using the $VAR[aX + b] = a^2 VAR[X]$ property of variance) $VAR[X] = VAR[2Y + 1] = 4VAR[Y]$. Also, Y is a Bernoulli random variable with $Y_i \in \{0, 1\}$ meaning Y^2 term is equal to Y. Therefore, $VAR[Y] = E[Y] - E[Y]^2 = p - p^2 = p(1 - p) = \frac{1}{2}(1 - \frac{1}{2}) = \frac{1}{4}$.

Therefore, the variance of X is

$$\mathbf{VAR}[\mathbf{X}] = VAR[2Y + 1] = 4VAR[Y] = 4 \cdot \frac{1}{4} = \mathbf{1}$$

Optionally, see the appendix for a long solution for variance using a conditional probability table and covariance.

Markov's Inequality can be applied since X is a non-negative random variable.

$$\begin{aligned} \textbf{Markov: } Pr[X \geq 3] &\leq \frac{E[X]}{t} \\ Pr[X \geq 3] &\leq \frac{2}{3} \end{aligned}$$

Chebyshev can be applied since the mean and variance are finite. However, the inequality must be rearranged since Chebyshev bounds are applied in terms of distance from the mean (in both positive and negative directions).

$$\begin{aligned} X &\geq 3 \\ X - 2 &\geq 3 - 2 \\ |X - 2| &\geq 1 \end{aligned}$$

Chebyshev defines variance-sensitive bounds on the probability of this distance from the mean.

$$\textbf{Chebyshev: } Pr[|X-\mu| \geq k] \leq \frac{VAR[X]}{k^2}$$
$$Pr[|X-2| \geq 1] \leq \frac{1^2}{1^2}$$
$$Pr[|X-2| \geq 1] \leq 1$$

Chernoff's bound for sum of independent Bernoulli random variables requires independence, yet $X = Z + Y$ where Z is dependent on Y. Therefore, Chernoff's bound cannot be applied.

2.2 Question 2b

$X = X_1 + \ldots + X_n$ represents a sum of indicator random variables with $Pr[X_i = 1] = \frac{1}{4} = p$. By linearity of expectation and the fact that an indicator random variable X_i is a Bernoulli random variable with $E[X_i] = p$,

$$\mathbf{E[X]} = E[X_1] + \ldots + E[X_n] = \frac{1}{4} + \ldots + \frac{1}{4} = \mathbf{\frac{n}{4}}$$

For variance, it is given that $X_i + \ldots X_n$ are independent which implies the covariance can be ignored and $VAR[X] = VAR[X_1] + \ldots + VAR[X_n]$. The variance of a Bernoulli random variable is $E[X^2] - E[X]^2$. Since $X \in \{0, 1\}$, X^2 is simply X. Therefore, $VAR[X_i] = E[X] - E[X]^2 = p - p^2 = p(1-p)$ and

$$\mathbf{VAR[X]} = VAR[X_1] + \ldots + VAR[X_n] = n \cdot p(1-p)$$

Substituting $p = \frac{1}{4}$

$$\mathbf{VAR[X]} = n \cdot \frac{1}{4}(1 - \frac{1}{4}) = \mathbf{\frac{3n}{16}}$$

For $t = \frac{3n}{10}$, **Markov**'s Inequality is applicable since X is non-negative.

$$\textbf{Markov: } Pr[X \geq t] \leq \frac{E[X]}{t}$$
$$Pr[X \geq \frac{3n}{10}] \leq \frac{\frac{n}{4}}{\frac{3n}{10}}$$
$$Pr[X \geq \frac{3n}{10}] \leq \frac{5}{6}$$

Chebyshev bounds can also be applied since the mean and variance are finite. Again, the inequality $Pr[X \geq \frac{3n}{10}]$ needs to be rewritten as the distance from the mean.

$$
\begin{aligned}
X &\geq \frac{3n}{10} \\
X - \frac{n}{4} &\geq \frac{3n}{10} - \frac{n}{4} \\
|X - \frac{n}{4}| &\geq \frac{3n}{10} - \frac{n}{4} \\
|X - \frac{n}{4}| &\geq \frac{n}{20}
\end{aligned}
$$

$$
\begin{aligned}
\textbf{Chebyshev: } Pr[|X - E[X]| \geq k] &\leq \frac{VAR[X]}{k^2} \\
Pr[|X - \frac{n}{4}| \geq \frac{n}{20}] &\leq \frac{\frac{3n}{16}}{(\frac{n}{20})^2} \\
Pr[|X - \frac{n}{4}| \geq \frac{n}{20}] &\leq \frac{\frac{3n}{16}}{\frac{n^2}{400}} \\
Pr[|X - \frac{n}{4}| \geq \frac{n}{20}] &\leq \frac{75}{n}
\end{aligned}
$$

Chernoff's bound can be applied since X is a sum of independent random Bernoulli variables.

Chernoff's bound also requires rearranging the inequality to calculate the bounds: $Pr[X > \frac{3n}{10}$ to $|X - \frac{n}{4} \geq \frac{n}{20}$.

$$
\begin{aligned}
\textbf{Chernoff: } Pr[|X - E[X]| \geq t] &\leq 2e^{-2\frac{t^2}{n}} \\
Pr[|X - \frac{n}{4}| \geq \frac{n}{20}] &\leq 2e^{-2\frac{(\frac{n}{20})^2}{n}} \\
&\leq 2e^{-2\frac{\frac{n^2}{400}}{n}} \\
&\leq 2e^{\frac{-2n}{400}} \\
&\leq 2e^{-\frac{n}{200}}
\end{aligned}
$$

3 Appendix

Z is a transformation of Y by an added constant of 1. Therefore, sample space is $\{1, 2\}$ and the probabilities are $Pr[Z = 1] = Pr[Z = 2] = \frac{1}{2}$. The expected value is

$$E[Z] = \sum_{j=1}^{2} p(z_j) \cdot z_j = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = \frac{3}{2}$$

	$Z=1$	$Z=2$
$Y=0$	$Pr[X=0+1=1]=\frac{1}{4}$	$Pr[X=0+2=2]=\frac{1}{4}$
$Y=1$	$Pr[X=1+1=2]=\frac{1}{4}$	$Pr[X=1+2=3]=\frac{1}{4}$

Table 1: X probability table

X is the sum of Y and Z. The sample space and probabilities of all possible X outcomes can be derived by joint probability of Y and Z.

Therefore, X is a discrete random variable with possible outcomes of 1,2, or 3. X can be represented as a sum of indicator variables with $Pr[X_i]$ defined by table 1, or the expectation can be computed by summing the expected value of Y and Z.

$$\begin{aligned}\mathbf{E}[\mathbf{X}] = \sum_{1}^{3} p(x_j)\cdot x_j = 1\cdot\frac{1}{4}+2\cdot\frac{1}{2}+3\cdot\frac{1}{4} = \frac{1}{4}+1+\frac{3}{4} &= \\ E[X_1]+E[X_2]+E[X_3] &= \\ E[Y+Z] &= \\ E[Y]+E[Z] &= \\ \frac{1}{2}+\frac{3}{2} &= \\ &\mathbf{2}\end{aligned}$$

For variance, it must be noted that Z depends on Y, so the $COVAR(X,Y)$ needs to be taken into account

$$VAR[X] = VAR[Z] + VAR[Y] + 2COVAR(Z,Y)$$

Y and Z are Bernoulli random variables with $p=\frac{1}{2}$. The variance of a random variable is $E[X^2]-E[X]^2$. Since Z and Y are Bernoulli random variables with $Z_i, Y_i \in \{0,1\}$, the X^2 term is equal to X. Therefore, $VAR[X_{Bernoulli}] = E[X]-E[X]^2 = p-p^2 = p(1-p)$

$$VAR[Y] = VAR[Z] = p(1-p) = \frac{1}{2}(1-\frac{1}{2}) = \frac{1}{4}$$

$E[ZY]$ is the weighted sum of the products of Z and Y.

$$E[ZY] = \frac{1}{2}(0\cdot 1)+\frac{1}{2}(1\cdot 2) = 1$$

Therefore, the covariance between Z and Y is

$$COVAR[Z,Y] = E[ZY] - (E[Z]\cdot E[Y]) = 1-(\frac{3}{2}\cdot\frac{1}{2}) = \frac{1}{4}$$

Finally, the variance of X can be calculated from the variance of Z, Y, and their covariance. Again, $VAR[Z] = VAR[Y] = \frac{1}{4}$; their the covariance is also $\frac{1}{4}$.

$$\mathbf{VAR[X]} = VAR[Z] + VAR[Y] + 2COVAR(Z,Y) = \frac{1}{4} + \frac{1}{4} + 2(\frac{1}{4}) = \mathbf{1}$$