1. **Ans.** We first consider the case where \( k \leq n/2 \). For this case, we scan the \( n \) items and obtain the minimum value, say, \( t \). Then, we add \( n - 2k + 1 \) items to the set of \( n \) items, each item having a value equal to \( t \). It is easy to check that the \( k \)th largest item among the original \( n \) items will be the median of the new set of \( 2n - 2k + 1 \) items. Therefore, we can apply the median-finding algorithm for the new set to obtain the desired \( k \)th largest item of the original set. The running time is \( O(n) \).

More precisely, we have spent \( n \) comparisons to find the minimum element, and at most \( 4n + o(n) \) further comparisons to find the median. The total time is thus at most \( 5n + o(n) \). In fact, it is easy to see that we do not need to explicitly add the \( n - 2k + 1 \) items. What we need to do is to modify the sampling algorithm a bit, so that with probability \( p = n/(2n - 2k + 1) \), we are selecting an integer from the original \( n \) items, and with probability \( 1 - p \), we are selecting the minimum value \( t \). Consequently, the time will be at most \( 3n + o(n) \).

For the case where \( k > n/2 \), we proceed by adding items with maximum value instead. The running time for this case is also \( O(n) \).

2. **Ans.** Firstly,

\[
\Pr \left( \left| \frac{X_1 + X_2 + X_3 + \ldots + X_n}{n} - \mu \right| > \varepsilon \right) = \Pr(\left| (X_1 + X_2 + \ldots + X_n) - n\mu \right| > n\varepsilon).
\]

By Chebyshev’s inequality and the independence of \( X_i \)'s, we have:

\[
\Pr(\left| (X_1 + X_2 + \ldots + X_n) - n\mu \right| > n\varepsilon) \leq \frac{\text{Var}[X_1 + X_2 + \ldots + X_n]}{(n\varepsilon)^2} = \frac{\sum_{i=1}^{n}\text{Var}[X_i]}{n^2\varepsilon^2} = \frac{n\sigma^2}{n^2\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.
\]

Combining, we have

\[
0 \leq \lim_{n \to \infty} \Pr \left( \left| \frac{X_1 + X_2 + X_3 + \ldots + X_n}{n} - \mu \right| > \varepsilon \right) \leq \lim_{n \to \infty} \frac{\sigma^2}{n\varepsilon^2} = 0
\]

which completes the proof of the weak law of large numbers.

3. **Ans.**

(a) Given \( X \sim \text{Bin}(n, p) \). Then, the MGF of \( X \), \( M_X(t) \), can be calculated as follows:

\[
M_X(t) = \mathbb{E}[e^{tX}] = \sum_{k=0}^{n} \Pr(X = k) e^{tk}
\]
\[
\begin{align*}
&= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{tk} \\
&= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k (1-p)^{n-k} \\
&= (pe^t + 1 - p)^n
\end{align*}
\]

(b) Since \(X\) and \(Y\) are independent, the MGF of \(X + Y\) is thus:
\[
M_{X+Y}(t) = M_X(t) \times M_Y(t) = (pe^t + 1 - p)^{n+m}.
\]

(c) The MGF in part (b) is the same as the MGF of a binomial random variable with parameters \(n + m\) and \(p\). Thus, \(X + Y \sim \text{Bin}(n + m, p)\).

4. (20%) Ans.

Let \(X\) be the effect of combined noise. We may express \(X\) by \(p_i\) and \(b_i\) as:
\[
X = p_1b_1 + p_2b_2 + ... + p_nb_n.
\]

Also, \(\mu = \mathbb{E}[X] = 0\).

From the question, we see that if the combined noise \(|X| < 1\), the signal can be decoded properly. Thus,
\[
\Pr(\text{receiver makes an error}) \leq \Pr(|X| \geq 1).
\]

Below, we shall give a Chernoff bound for \(\Pr(|X| \geq 1)\). First, we calculate the moment generating function of each \(p_i b_i\):
\[
M_{p_i b_i}(t) = \mathbb{E}[e^{tp_i b_i}] = \frac{1}{2} \left( e^{tp_i} + e^{-tp_i} \right)
= \sum_{k \geq 0} \frac{(tp_i)^{2k}}{(2k)!}
\leq e^{t^2 p_i^2 / 2}.
\]

This implies
\[
M_X(t) \leq e^{t^2 K / 2}, \quad \text{where } K = \sum_i p_i^2.
\]

Thus, we have:
\[
\begin{align*}
\Pr(|X| \geq 1) &= 2 \Pr(X \geq 1) \quad \text{(X is symmetric around mean } \mu = 0) \\
&= 2 \Pr(e^{tX} \geq e^t) \\
&\leq \frac{2 \mathbb{E}[e^{tX}]}{e^t} = \frac{2M_X(t)}{e^t} \\
&\leq \frac{2e^{t^2 K / 2}}{e^t}.
\end{align*}
\]

The last term is minimized when we set \(t = 1/K\), so that we obtain:
\[
\Pr(|X| \geq 1) \leq 2e^{-1/(2K)}.
\]
5. Ans.

(a) The expected load of each server is:
\[
\bar{L} = \frac{1}{m} \sum_{i=1}^{n} L_i.
\]

(b) Let \(R_j\) denote the load assigned to server \(j\). Then we have \(E[R_j] = \bar{L}\).
On the other hand, let \(R_{j,i}\) be a random variable such that it has value \(L_i\) if job \(J_i\) is
assigned to server \(j\), and it has value 0 otherwise. Then, we see that:
\[
R_j = R_{j,1} + R_{j,2} + \cdots + R_{j,n}.
\]
Since \(R_j\) is the sum of independent random variables whose range is between 0 and 1, we can apply
Theorem 1 immediately, and obtain:
\[
\Pr(R_j \geq c\bar{L}) = \Pr(R_j \geq cE[R_j]) \leq e^{-\alpha E[R_j]} = e^{-\alpha \bar{L}},
\]
where \(\alpha = c \ln c + 1 - c\).
Thus by union bound,
\[
\Pr(\text{some server has load at least } c\bar{L}) \leq me^{-\alpha \bar{L}}.
\]

(c) Given \(n = 100K\), \(m = 10\), and the average job execution time is 0.25 sec. So we have:
\[
\sum_{i=1}^{n} L_i = 0.25n = 25K \quad \text{and} \quad \bar{L} = 2500.
\]
We hope to bound
\[
\Pr(\text{all servers have load at most } 1.1\bar{L})
\]
so that we set \(c = 1.1\), and consequently \(\alpha\) is about 0.004841.
By the result in part (b), we see that
\[
\Pr(\text{some server has load at least } 1.1\bar{L}) \leq 10 e^{-0.004841 \times 2500} \leq e^{-9}.
\]
Thus, we get the desired bound that
\[
\Pr(\text{all servers have load at most } 1.1\bar{L}) \geq 1 - e^{-9}.
\]

6. (Bonus: 10%) 

(a) Let \(f(z) = c^{-\lambda}(1 + \lambda(c - 1)) + z(c^{1-\lambda} - c^{-\lambda}) - c^z\). So we get
\[
f'(z) = c^{1-\lambda} - c^{-\lambda} - \ln c \cdot c^z \quad \text{and} \quad f''(z) = -(\ln c)^2 \cdot c^z.
\]
Observe that \(f''(z) < 0\), which implies \(f'(z)\) is strictly decreasing. Then we see that:
\[
\begin{align*}
  f'(-\lambda) &= c^{1-\lambda} - c^{-\lambda} - \ln c \cdot c^{-\lambda} \\
  &= c^{1-\lambda} - (1 + \ln c) \cdot c^{-\lambda} \\
  &> c^{1-\lambda} - (c) \cdot c^{-\lambda} \quad \text{(since } 1 + \ln c < e^{\ln c} = c) \\
  &= 0 \\
\end{align*}
\]

\[
\begin{align*}
  f'(1 - \lambda) &= c^{1-\lambda} - c^{-\lambda} - \ln c \cdot c^{1-\lambda} \\
  &= -c^{-\lambda} + (1 - \ln c) \cdot c^{1-\lambda} \\
  &< -c^{-\lambda} + (1/c) \cdot c^{1-\lambda} \quad \text{(since } 1 - \ln c < e^{-\ln c} = 1/c) \\
  &= 0.
\end{align*}
\]

As \( f'(z) \) is continuous and strictly decreasing, the above statements indicates that \( f'(z) = 0 \) occurs when \( z \) is in \([-\lambda, 1 - \lambda] \), so that the global maximum of \( f(z) \) is attained at the corresponding value of \( z \).

Thus, it must be true that:

\[
f(z) \geq \min\{f(-\lambda), f(1 - \lambda)\} = 0.
\]

Thus, \( c^z \leq c^{-\lambda}(1 + \lambda(c - 1)) + z(c^{1-\lambda} - c^{-\lambda}) \), so that

\[
c^z \leq c^{-\lambda}e^{\lambda(c-1)} + z(c^{1-\lambda} - c^{-\lambda}).
\]

(b)

\[
\Pr(X \geq cE[X]) = \Pr\left( \sum_i X_i \geq cE[X] \right)
\]

\[
= \Pr\left( \sum_i (\Delta_i + E[X_i]) \geq cE[X] \right)
\]

\[
= \Pr(\Delta + E[X] \geq cE[X])
\]

\[
= \Pr(\Delta \geq (c - 1)E[X])
\]

\[
= \Pr(c\Delta \geq c^{(c-1)E[X]})
\]

\[
\leq \frac{E[c\Delta]}{c^{(c-1)E[X]}}.
\]

(c) By setting \( \lambda = E[X_i], \)

\[
E[e^{c\Delta_i}] \leq E[e^{-\lambda}e^{\lambda(c-1)} + \Delta_i(c^{1-\lambda} - c^{-\lambda})]
\]

\[
= E[e^{-\lambda}e^{\lambda(c-1)}] \quad \text{(since } E[\Delta_i] = E[X_i] - E[X_i] = 0) \\
= e^{-\lambda}e^{\lambda(c-1)} \quad \text{(since } \lambda = E[X_i] \text{ is a constant)}
\]

\[
= e^{\lambda(c-1-\ln c)}.
\]

(d) Since \( \Delta_1, \Delta_2, \ldots, \Delta_n \) are independent, we have:

\[
E[c\Delta] = \prod E[e^{\Delta_i}] \leq \prod e^{E[X_i](c-1-\ln c)} = e^{E[X](c-1-\ln c)}.
\]
Combining this with part (b), we have:

$$\Pr(X \geq c E[X]) \leq \frac{e^{E[X](c-1-\ln c)}}{c^{(c-1)E[X]}}$$

$$= \frac{e^{E[X](c-1-\ln c)}}{e^{(c \ln c - \ln c)E[X]}}$$

$$= e^{E[X](c-1-\ln c-c \ln c+\ln c)}$$

$$= e^{-\alpha E[X]},$$

where $\alpha = c \ln c + 1 - c$. 