Question 1:

For size m and length n. suppose we only have one hash function:

then the probability of some targets not hit by a dart is: $(1-\frac{1}{n})^m$

so on the contrary, the prob that the target hit by at least one dart is:

$$1-(1-\frac{1}{n})^m = 1-(1-\frac{1}{n})^{n\cdot\frac{m}{n}}$$

when the data's size goes huge, set $n\to\infty$, there is: $1-(1-\frac{1}{n})^{n\cdot\frac{m}{n}} = 1-e^{-\frac{m}{n}}$

so when we have k hash functions, the probability goes to: $1-e^{-k\frac{m}{n}}$

and due to the definition, the false positive prob is: $(1-e^{-k\frac{m}{n}})^k$

set $f(k)=(1-e^{-k\frac{m}{n}})^k$ use some calculation applications, the minimum is $k'=\ln 2\cdot\frac{n}{m}$

and sometime k is not an integer.

so $k^* = \operatorname{argmin}(f(\lfloor k^* \rfloor), f(\lfloor k^* \rfloor + 1))$

Question 2:

Use the definition: $M|_k = \sum_{i \in A} (m_i)^k$

where k is the moment

m_i is the number of times of value i occurs in the stream

so $m_1 = 3,\ m_2 = 2,\ m_3 = 2,\ m_4 = 2$

$M|_2 = \sum_{i \in A} (m_i)^2 = 3^2 + 2^2 + 2^2 + 2^2 = 21$

$M|_3 = \sum_{i \in A} (m_i)^3 = 3^3 + 2^3 + 2^3 + 2^3 = 51$

Question 3:

(a): Estimate the average purchase price

Key attribute: Item purchased

Approach: Randomly select $\frac{1}{20}$ th of the transactions for each item purchased, and calculate the average purchase price for each selected group.

Then estimate the average purchase price for each item by taking the average of the averages from all selected groups

(b) Estimate the fraction of custermors who made a purchase of $50 or more

Key attribute: Custormer ID

Approach: Randomly select $\frac{1}{20}$ th of the transactions for each customer and count the number of transactions where the purchase price was $50 or more. Finally estimate the fraction of customers who made a purchase of $50 or more by taking the average of the fractions from all selected groups.

(c): Estimate the fraction of items that were purchased by at least 10 customers

Key attribute: Item purchased

Approach: Randomly select 1/20 th of the transactions for each item purchased and count taking the number of transactions. Finally, estimate the fraction of items that were purchased by the average of the fractions from all selected groups