a). Ex 6.10 from FOS.

Let n be the number of flips needed to see the 1st head.

$P(n=1) = p$, $P(n=2) = (1-p)(p)$, $P(n=3) = (1-p)^2(p)$ $\cdots$

Expectation:

$$E(n) = \sum_{k=1}^{\infty} k \cdot P(n=k)$$

$$= \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} p$$

$$= \frac{p}{1-p} \sum_{k=1}^{\infty} k \cdot (1-p)^{k} \quad , \quad \text{let } (1-p) = r$$

$$= \frac{p}{1-p} r \sum_{k=0}^{\infty} k \cdot r^{k-1} \quad , \quad \text{we have } \frac{d}{dr} r^k = k \cdot r^{k-1}$$

$$= \frac{p}{1-p} r \frac{d}{dr} \left(\sum_{k=0}^{\infty} r^k \right) \quad , \quad \text{since } |r| < 1$$

$$= \frac{p}{1-p} r \frac{d}{dr} \left(\frac{1}{1-r} \right)$$

$$= \frac{p}{1-p} r \frac{1}{(1-r)^2}$$

$$= \frac{p}{1-p} \cdot (1-p) \cdot \frac{1}{p^2}$$

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

b) Ex.2.3.1 MMDS

(a) The lagest number.

Map: ~~f~~ map all number to same key, $\langle max, n_i \rangle$, n_i is the number

Reduce: then we have only one key-value pair : $\langle max, [n_1, n_2, \cdots, n_i] \rangle$, reduce it by selecting the maximum of value list. The output is $\langle max, n_{max} \rangle$, n_{max} is the max for current machine.

Map: map all the outputs of last stage to themselves (identity).

Reduce: we have $\langle max, [n_{max1}, n_{max2}, \cdots] \rangle$, where n_{maxi} represents the maximum of different distributed machine, reduce it by selecting the maximum of value list. The output is

$\langle max, n_{MAX} \rangle$, n_{MAX} is the largest number.

(b) The mean of data set.

Map: same as last question : $\langle mean, n_i \rangle$

Reduce: compute the mean of value list : $\langle mean, \frac{n_1+n_2+\cdots n_i}{i} \rangle$

Map: Identity

Reduce: $\langle mean, [\mu_1, \mu_2, \cdots, \mu_k] \rangle \rightarrow \langle mean, \frac{\mu_1+\mu_2+\cdots+\mu_k}{k} \rangle$

(c) The set of integer.

Map: map each number n_i to a key-value pair ~~$\langle n_i, 1 \rangle$~~ $\langle set, n_i \rangle$

Reduce: reduce $\langle set, [n_1, n_2, \cdots] \rangle$ to $\langle set, set([n_1, n_2, \cdots]) \rangle$

Map : Identity

Reduce: $\langle set, [set_1, set_2, \cdots] \rangle \rightarrow \langle set, set([set_1, set_2, \cdots]) \rangle$

~~(d) The count of distinct numbers.~~

~~Map: map each number to $\langle n_i, 1 \rangle$~~

~~Reduce: $\langle number, [1, 1, \cdots] \rangle \rightarrow \langle number, 1 \rangle$~~

(d) The count of distinct numbet.

~~We just need to take the result of (c) problem~~

Map: map each number n_i to $\langle n_i, 1 \rangle$

Reduce: remove duplicated 1's for each key, so that each distinct number will have a key-value pair $\langle num, 1 \rangle$

Map: identity

Reduce: remove duplicated 1's for each key.

finally, we count the number of keys, that's exactly the same number of distinct number in data set.