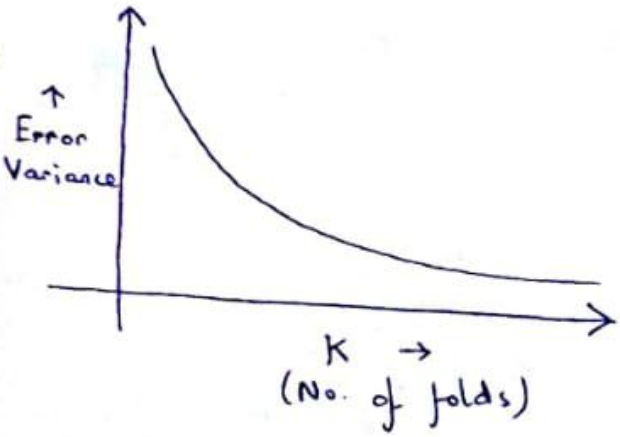
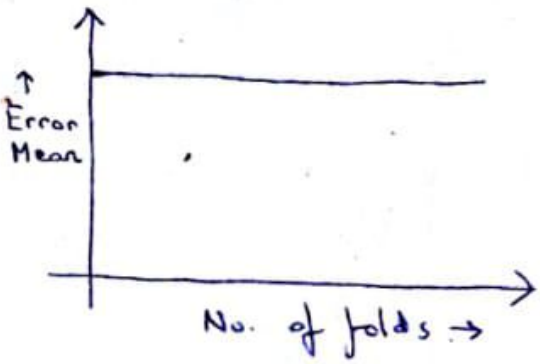
SMAI HW-6 Problem 1.

1 a) It is expected that the variance of the error is inversely proportional to k. i.e with increasing k, variance reduces.

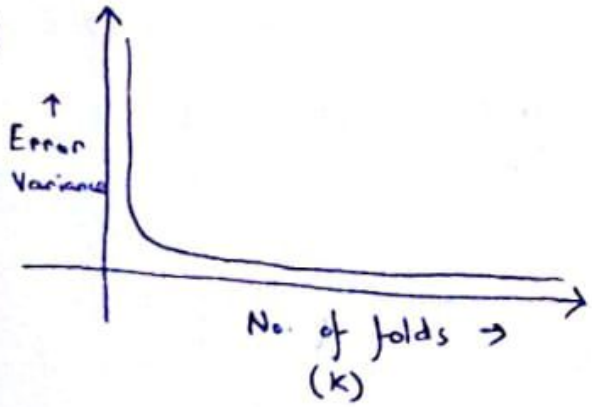
In contrast, the average mean is a constant, no matter what the value of k is. (k should divide the number of samples in population).

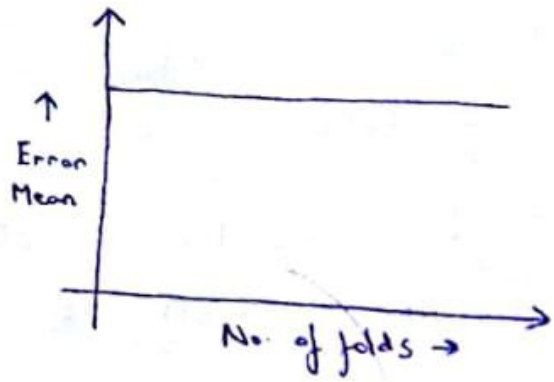
i)





ii)





For small values of k:

→ Small set of points is sampled.

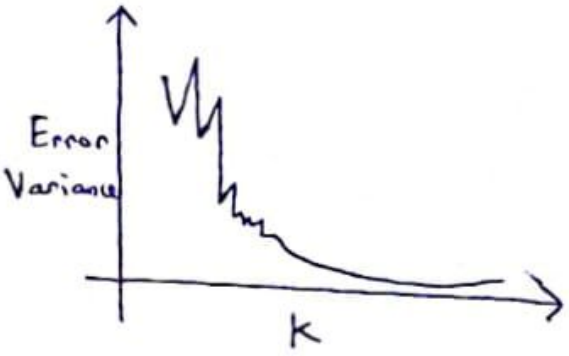
→ ~~Means~~ Their means are used to calculate error variance.

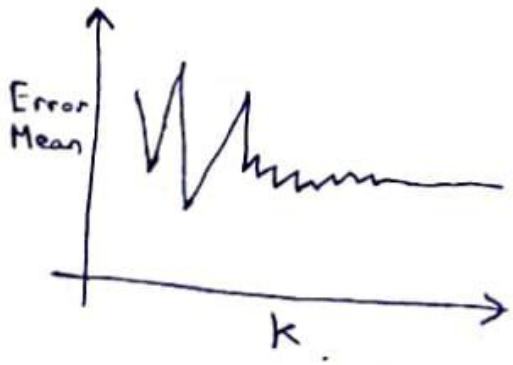
∴ We have large number of values from population of small sizes.

As K increases, large chunks of population get sampled, the number of chunks reduces; and we achieve a better representation of population.

Hence, the sample mean of each chunk tends to the population mean thereby reducing the variance betw each chunk. The mean error remains constant since the entire data's representation tends to that of the population's distribution.

If we try all values of K, we get following curves. for mean and variance:





→ This is due to uneven distribution of data points.

→ The variation will be more prominent for small K values.