

## Chapter 5 - Resampling Methods

1. Using  $Var(A) = E[A^2] - (E[A])^2$  we get,

$$F = Var(\alpha X + (1 - \alpha)Y) = E[(\alpha X + (1 - \alpha)Y)^2] - (E[\alpha X + (1 - \alpha)Y])^2$$

$$E[(\alpha X + (1 - \alpha)Y)^2] = E[\alpha^2 X^2 + (1 - \alpha)^2 Y^2 + 2\alpha(1 - \alpha)XY]$$

$$E[(\alpha X + (1 - \alpha)Y)^2] = \alpha^2 E[X^2] + (1 - \alpha)^2 E[Y^2] + 2\alpha(1 - \alpha)E[XY]$$

$$(E[\alpha X + (1 - \alpha)Y])^2 = (\alpha E[X] + (1 - \alpha)E[Y])^2$$

$$(E[\alpha X + (1 - \alpha)Y])^2 = \alpha^2 (E[X])^2 + (1 - \alpha)^2 (E[Y])^2 + 2\alpha(1 - \alpha)E[X]E[Y]$$

$$F = \alpha^2 (E[X^2] - (E[X])^2) + (1 - \alpha)^2 (E[Y^2] - (E[Y])^2) + 2\alpha(1 - \alpha)(E[XY] - E[X]E[Y])$$

$$F = \alpha^2 \sigma_X^2 + (1 - \alpha)^2 \sigma_Y^2 + 2\alpha(1 - \alpha)\sigma_{XY}$$

$$F = \alpha^2 [\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}] + 2\alpha[\sigma_{XY} - \sigma_Y^2] + \sigma_Y^2$$

To minimize  $\alpha$ , we set  $\frac{\partial F}{\partial \alpha} = 0$ .

$$\frac{\partial F}{\partial \alpha} = 2\alpha[\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}] + 2[\sigma_{XY} - \sigma_Y^2] = 0$$

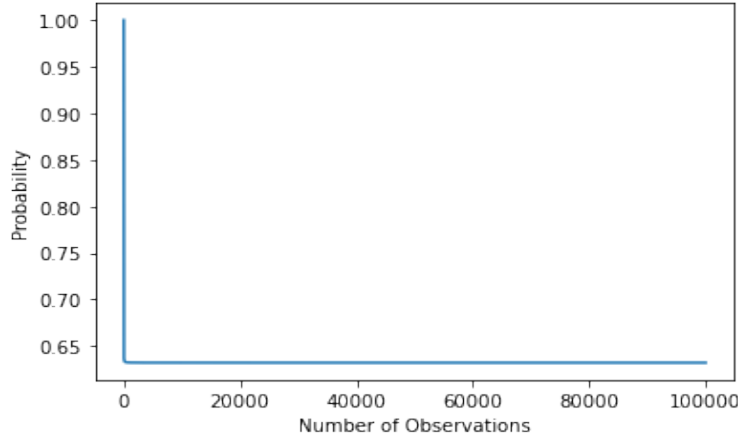
$$\text{Thus, } \alpha = \frac{\sigma_Y^2 - \sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}$$

2. (a) There are  $n$  observations, and we need to find the probability that the first observation is not the same as the  $j^{th}$  observation. This is the same as 1 - the probability that the first observation is the same as the  $j^{th}$  observation. Since each observation has a equal chance of being picked,  $P(O_1 = O_j) = \frac{1}{n}$ . Thus,  $P(O_1 \neq O_j) = 1 - \frac{1}{n}$
- (b) By the same reasoning as 2a.  $P(O_2 \neq O_j) = 1 - \frac{1}{n}$
- (c) Probability that the  $j^{th}$  observation is not in the bootstrap =  $\prod_{i=1}^n P(O_i \neq O_j) = (1 - \frac{1}{n})^n$
- (d) Probability that the  $j^{th}$  observation is in the bootstrap for  $n = 5$   $= 1 - \prod_{i=1}^5 P(O_i \neq O_j) = 1 - (1 - \frac{1}{5})^5$

- (e) Probability that the  $j^{th}$  observation is in the bootstrap for  $n = 100$   

$$= 1 - \prod_{i=1}^{100} P(O_i \neq O_j) = 1 - \left(1 - \frac{1}{100}\right)^{100}$$
- (f) Probability that the  $j^{th}$  observation is in the bootstrap for  $n = 10000$   

$$= 1 - \prod_{i=1}^{10000} P(O_i \neq O_j) = 1 - \left(1 - \frac{1}{10000}\right)^{10000}$$
- (g) Plotting the probability that the  $j^{th}$  observation is in the bootstrap for  $n = 1$  to 100000



3. (a) The training data set is broken into two sections of sizes  $\frac{n}{k}$  and  $\frac{n(k-1)}{k}$ . The larger set is used for training the model and the smaller set is used for validation. The evaluation metric is computed for the performance over the validation set. This is done  $k$  times, each time with a new combination for the validation and training sets. The evaluation metric is averaged out for all  $k$  instances. This average is the final cross validation performance metric.
- (b) i. K-fold cross validation uses, on average, more observations for training ( $\frac{n(k-1)}{k}$ ) as compared to the validation set approach which only uses half the number of observations. This leads to higher bias in the validation set approach. However, k-fold cross validation is computationally more expensive than the validation set approach.

- ii. K-fold cross validation uses, on average, lesser observations for training ( $\frac{n(k-1)}{k}$ ) as compared to the LOOCV which uses  $n - 1$  observations for every training cycle. This leads to very low bias but very high variance. K-fold cross validation is also computationally less expensive than LOOCV.
4. To estimate the standard deviation of the estimate we use the bootstrap method. In this method we reshuffle and reorder the data set to create a new data set. That is, this data set can have repeated observations. The model is trained on this data set and a prediction for  $Y$  is made. This process is repeated multiple times and we obtain multiple values for the prediction of  $Y$ . We can now find the standard deviation for these values and this would be our estimate for the standard deviation of the prediction of  $Y$ .