$$Var(aX+bY) = a^2 Var(X) + 2ab\,Cov(X,Y) + b^2 Var(Y)$$

Ch 5 #1

$$\alpha = \frac{\sigma_y^2 - \sigma_{xy}}{\sigma_y^2 + \sigma_x^2 - 2\sigma_{xy}}, \quad \sigma_x^2 = Var(X), \quad \sigma_y^2 = Var(Y), \quad \sigma_{xy} = Cov(X,Y)$$

Show $\alpha$ minimizes

$$Var(\alpha X + (1-\alpha)Y)$$

$$= \alpha^2\sigma_x^2 + (1-\alpha)^2\sigma_y^2 + 2\alpha(1-\alpha)\sigma_{xy}$$

$$\frac{\partial}{\partial\alpha} = \left[\alpha^2\sigma_x^2 + (1 - 2\alpha + \alpha^2)\sigma_y^2 + (2\alpha - 2\alpha^2)\sigma_{xy}\right]$$

$$= 2\alpha\sigma_x^2 + -2\sigma_y^2 + 2\alpha\sigma_y^2 + 2\sigma_{xy} - 4\alpha\sigma_{xy} = 0$$

Solving for $\alpha$

$$2\alpha\sigma_x^2 + 2\alpha\sigma_y^2 - 4\alpha\sigma_{xy} = 2\sigma_y^2 - 2\sigma_{xy}$$

$$2\alpha(\sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}) = 2\sigma_y^2 - 2\sigma_{xy}$$

$$\alpha = \frac{\sigma_y^2 - \sigma_{xy}}{\sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}}$$ ✓

So this indeed is a minimum

$$\frac{\partial^2}{\partial\alpha^2}\left(Var(\alpha X + (1-\alpha)Y)\right)$$

$$= 2\sigma_x^2 + 2\sigma_y^2 - 4\sigma_{xy} = 0$$

$$= 2(\sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}) = 0$$

$$Var(X) + Var(Y) - Cov(X,Y) \geq 0$$

$$2Var(X-Y)$$

2 a) Probability that the first bootstrap obs is not the jth obs

$$1 - \frac{1}{n}$$

b) $1 - \frac{1}{n}$ for second obs

c) this is because the bootstrap sample (w/ replacement) is the product of the observations

so $\left(1 - \frac{1}{n}\right)_1 \left(1 - \frac{1}{n}\right)_2 \cdots \left(1 - \frac{1}{n}\right)_n$

$$= \left(1 - \frac{1}{n}\right)^n$$

d) when $n = 5$ $P(5\text{th obs}) = ?$ in Bootstrap sample

$$\Rightarrow 1 - \left(1 - \frac{1}{5}\right)^5 = 1 - (.8)^5 = 1 - .32768$$

$$= .67232 \approx .672$$

e) How about $n = 100$?

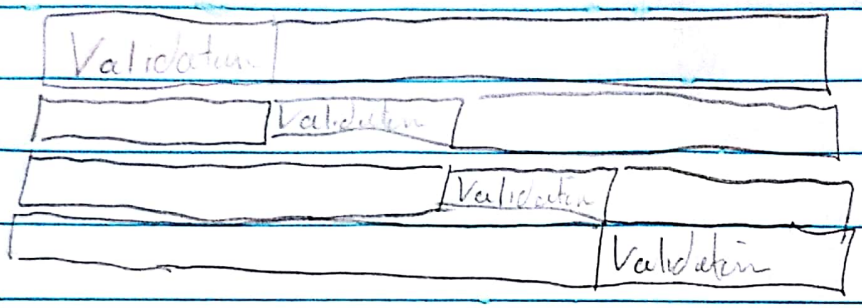
$$1 - \left(1 - \frac{1}{100}\right)^{100} = 1 - (.99)^{100} = .634$$

f) $n = 1000$?

$$1 - \left(1 - \frac{1}{1000}\right)^{1000} = .6323$$

g) R Code

K-fold CV is implemented by randomly dividing the observations into K groups. Each fold has a Validation & training set that are used to estimate the MSE for that fold. This is repeated for each fold




K=4 CV

b) What is advantages/disadvantages of kfold CV

i.) Compared to the Validation Set approach

In the Validation set approach there is alot of variance compared to kfoldCV so kfoldCV has the advantage

ii) Compared to LOOCV

LOOCV cost more time to calculate compared to k-fold CV, kfold CV has far less training fit time/iterations. Also LOOCV may given a Biase estimate because of the overfitting (High correlation/ training sets)