4.4

Intro

In this exercise, we will proof that when we have a linear relationship between our r.v's the correlation coefficient is equal to -1/+1. This makes sense because the correlation coefficient measures the 'degree of linear dependence' between r.v's. Thus, when the coefficient reach its extreme values, we should expect a full linear relationship between the variables.

Solution

Let's not make any assumption about the sign of a for now:

$$Cov(X, Y) = E[XY] - E[X]E[Y] \tag{1}$$

Let's substittue $Y = aX + b$ in Equation 1:

$$\begin{aligned} &E[XY] - E[X]E[Y] = E[aX^2 + bX] - E[X](aE[X] + b) = \\ &aE[X^2] + bE[X] - aE[X]^2 - bE[X] = a(E[X^2] - E[X]^2) = aVar(X) \end{aligned} \tag{2}$$

Using Equation 2 in the definition of the correlation coefficient, we arrive at:

$$\begin{aligned} &\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} = \frac{aVar(X)}{\sqrt{a^2Var(X)^2}} = \\ &\frac{aVar(X)}{|a|Var(X)} = sign(a) \end{aligned} \tag{3}$$

where $sign(a)$ is the sign function. Therefore, $a > 0 \implies \rho = 1$ and $a < 0 \implies \rho = -1$

Conclusion

In this exercise, we proved that r.v's that have linear relationship have the minimum/maximum correlation coefficent. As mentioned in the introduction, this makes a lot of sense, because $\rho(X, Y)$ is a measure of the linear dependence between X and Y.