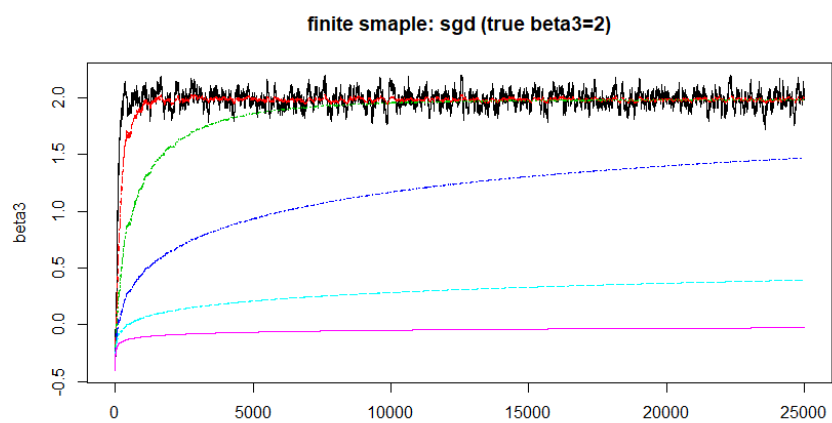
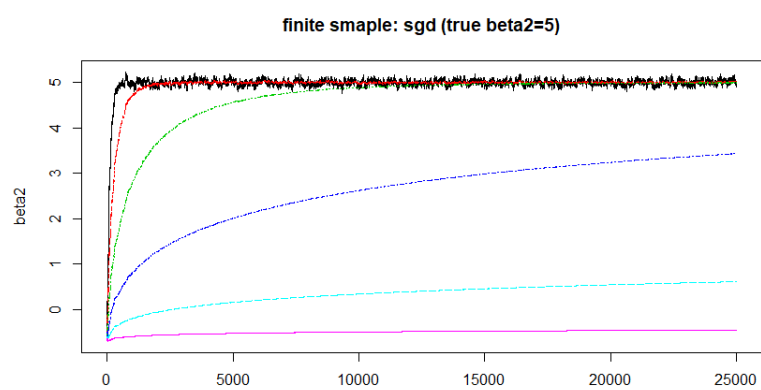
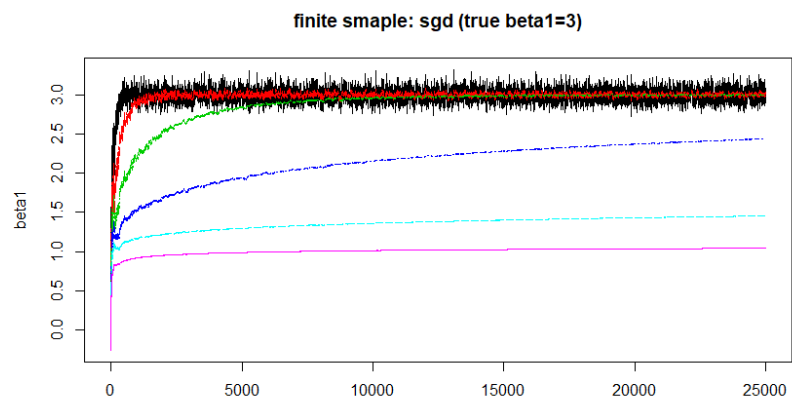
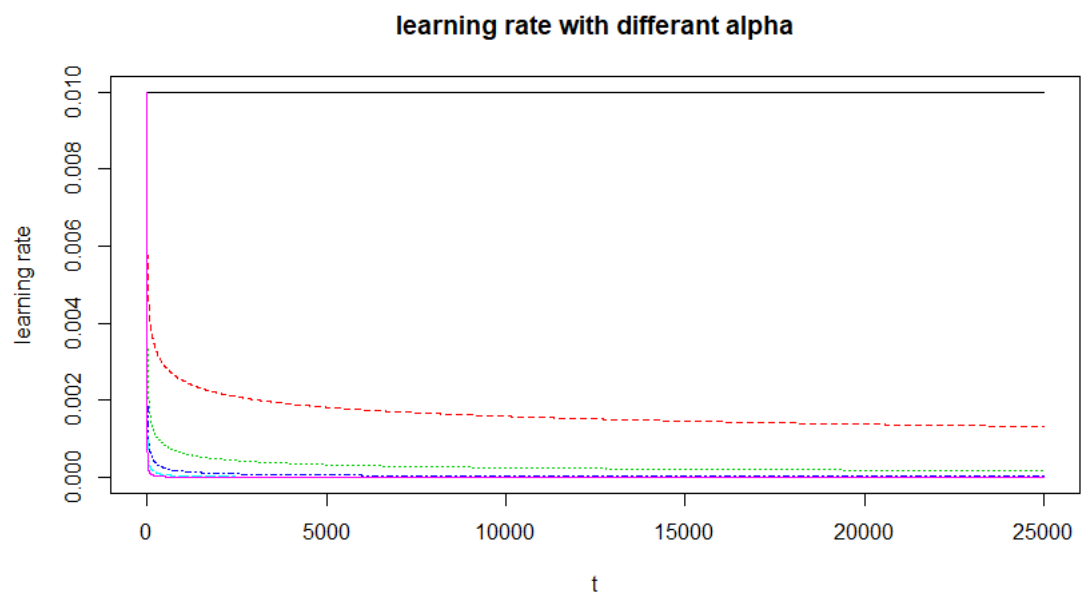


以下的每個圖，線條由上到下表示 α 越大(0.0,0.2,0.4,0.6,0.8,1)，可以發現當 α 越大， β 估計值會越快停止變化。

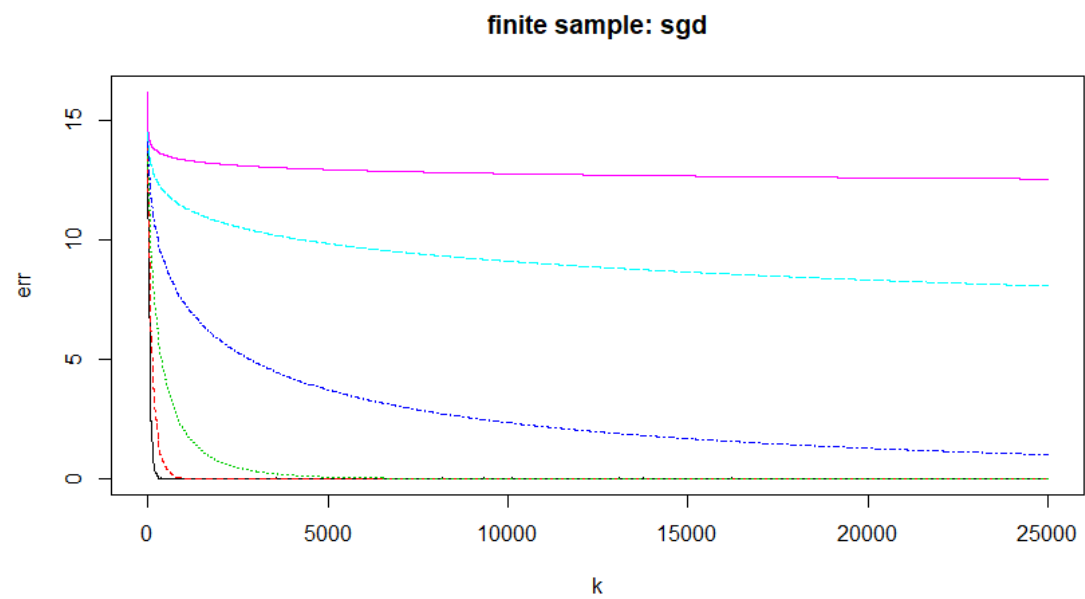


不同 α 下的學習率變化:

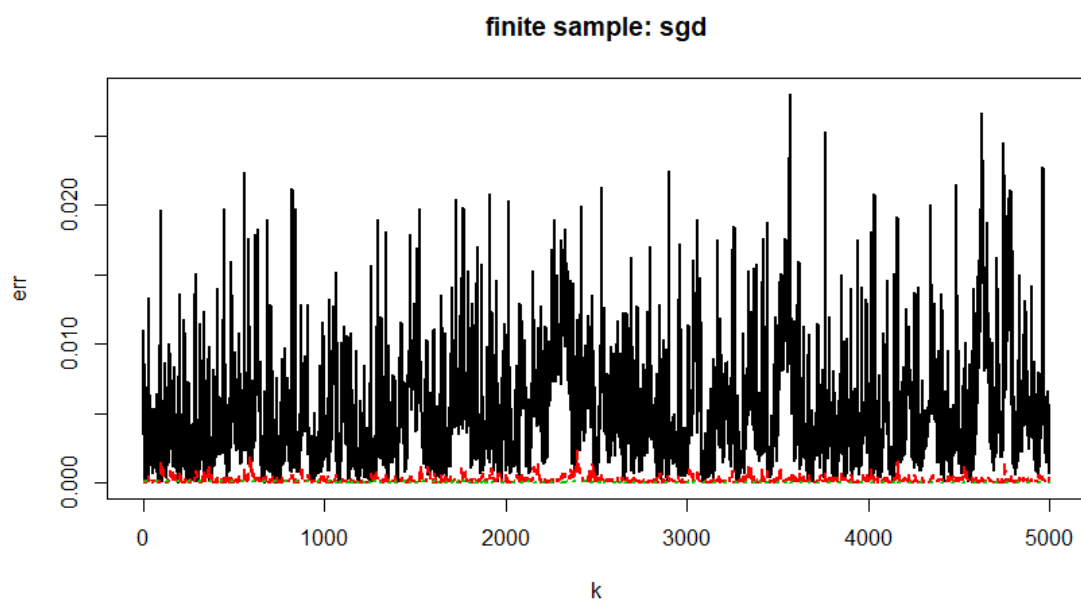


誤差大小:

看似固定常數(黑色線條)可以獲得最好的結果，但若看後面 5000 筆作比較。

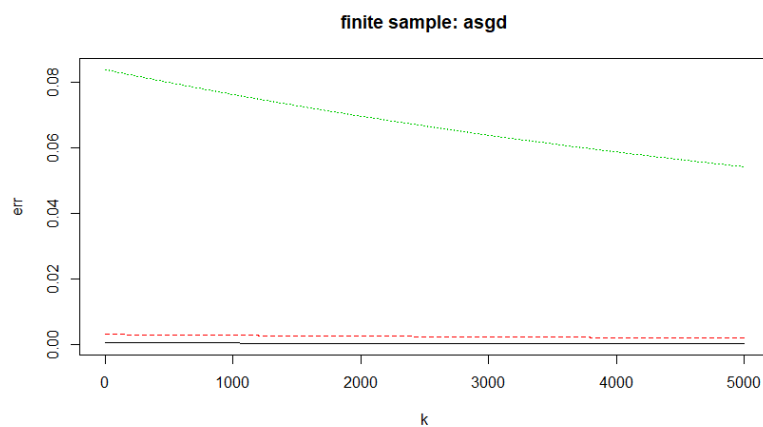
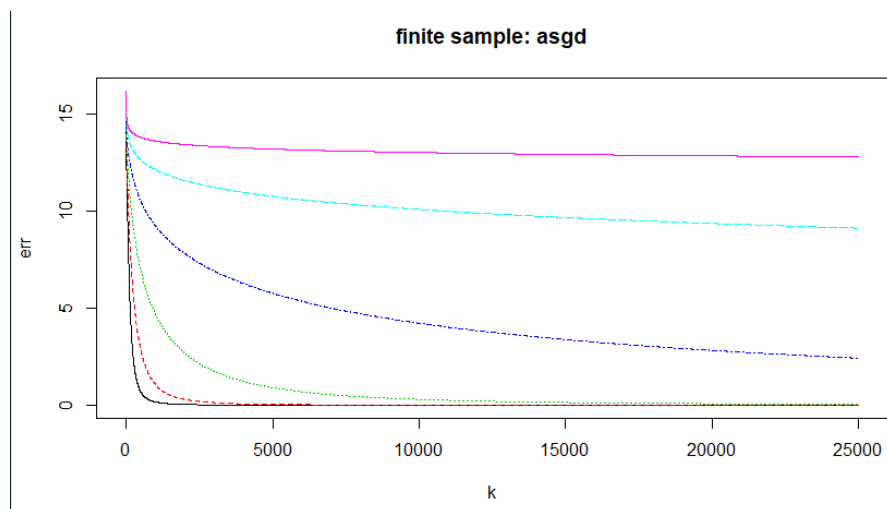


在 $\alpha=0.2$ (紅) 及 $\alpha=0.4$ (綠)的情況下 SGD 能獲得最小誤差



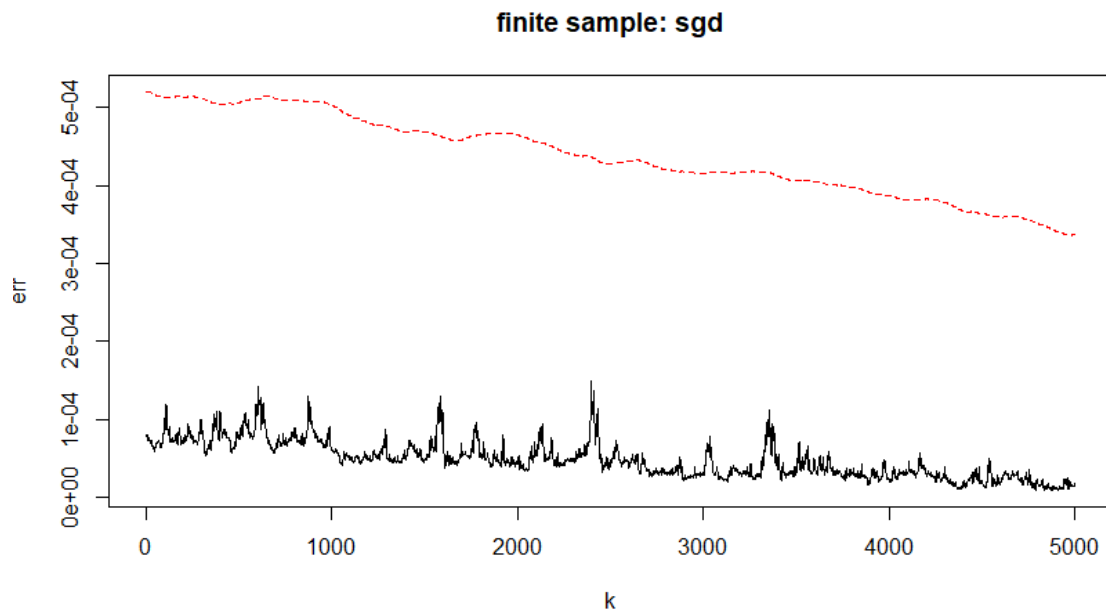
另一方面:

ASGD 則是固定學習率獲得最好的表現



比較 SGD 收斂最好的參數及 ASGD 收斂最好的參數:

SGD 在 $\alpha=0.4$ 的情況下與 ASGD 在 α 固定為 0 的情況下比較發現，SGD 的誤差較小一些。

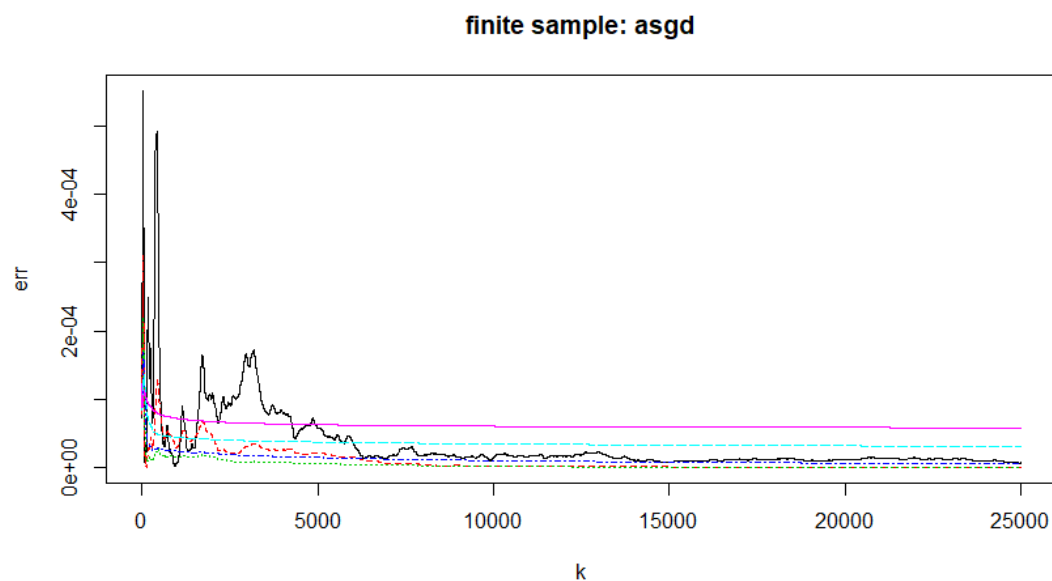


為何 ASGD 進行 γ_{decay} 後結果較差?

→ 可能原因: 估計值太晚跳動至目標參數導致前半段的誤差影響大

→ 將真實參數設定為起始值進行比較

→ 可以發現，當 α 太大導致估計值基本上不存在變化時，誤差也跟著固定住了，但是若還是存在著些微變化，其效果可以優於固定常數。



且將固定起始值後得到的 SGD 及 ASGD 最佳參數做比較，圖中黑色為 SGD($\alpha=0.6$)與紅色 ASGD($\alpha=0.4$)的情況下，ASGD 會擁有較小的誤差

