Exercise 4.8 *(programming)*   Value iteration. Let us first replicate the solution of the Example 4.3, where $p_h = 0.4$. After 32 iterations of the value-iteration algorithm, the value estimates are fully converged (see Fig. 1). The entire family of optimal policies found by the argmax action selection with respect to the optimal value function is depicted in Fig. 2. Next, in the case of $p_h = 0.25$, the family of the optimal policies is identical to the previous case (Fig. 2), whereas the value estimates look slightly different, see Fig. 3. The situation is completely different in the case of $p_h = 0.55$; see the corresponding value function and optimal policies in Fig. 4 and Fig. 5, respectively. In this case, the found family of the optimal policies consists of stakes of zero and one for each value of the gambler's capital.
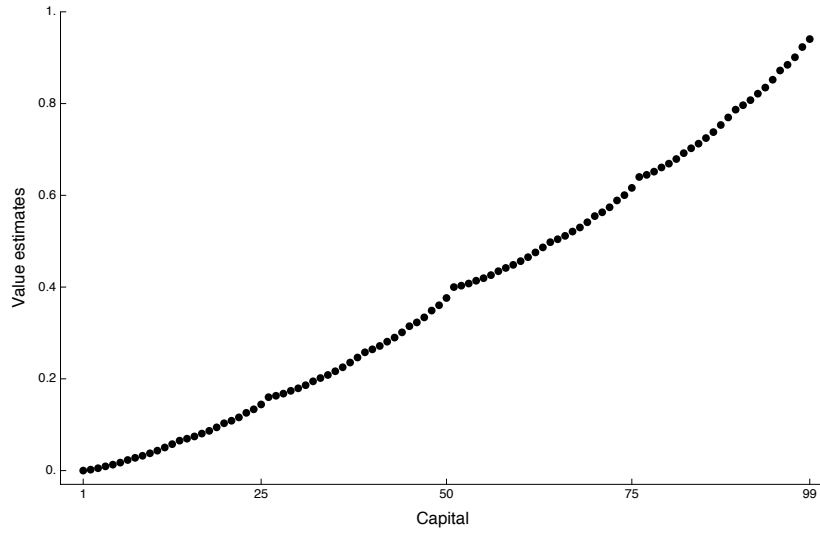

Figure 1: *Example 4.3*: Final value function for the Gambler's Problem for $p_h = 0.4$.
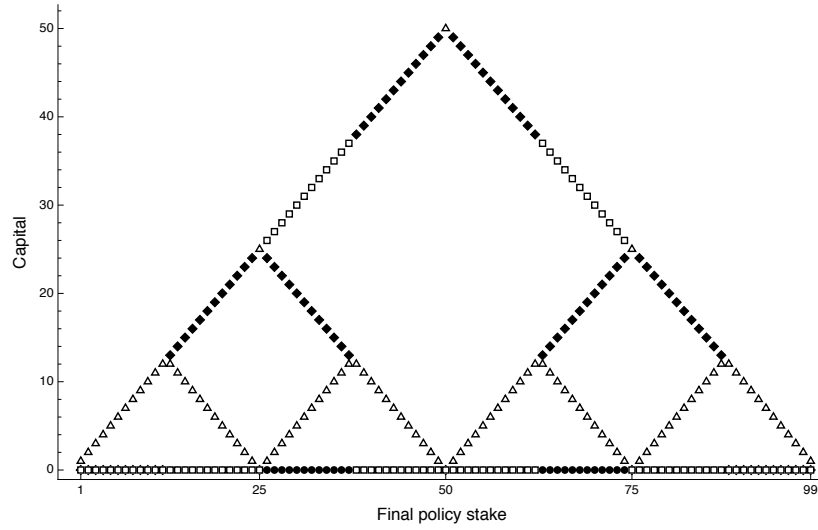
Figure 2: *Example 4.3 and Exercise 4.5*: Entire family of optimal policies for the Gambler's Problem for $p_h = 0.4$ and $p_h = 0.25$ (the two families are identical).
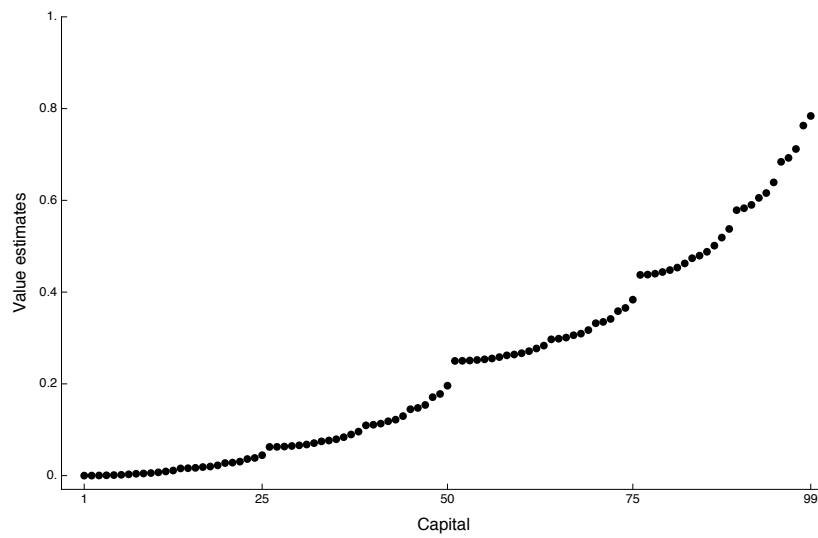

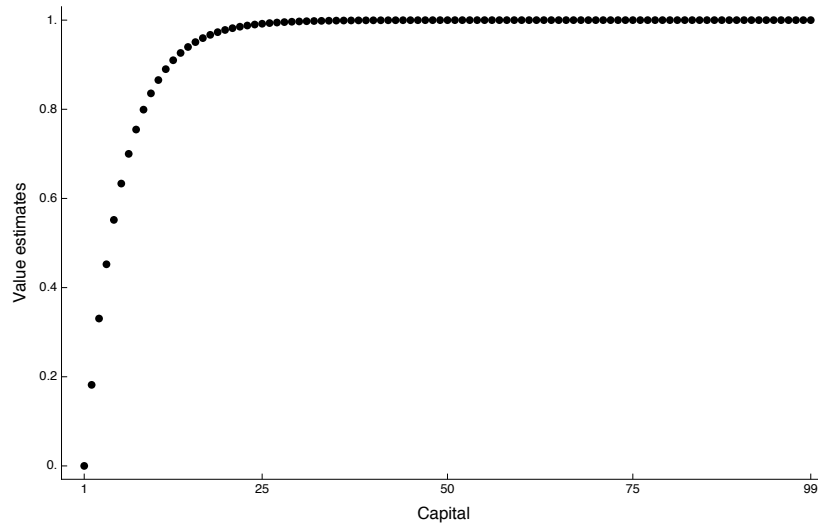Figure 3: *Exercise 4.5*: Final value function for the Gambler's Problem for $p_h = 0.25$.

Figure 4: *Exercise 4.5*: Final value function for the Gambler's Problem for $p_h = 0.55$.
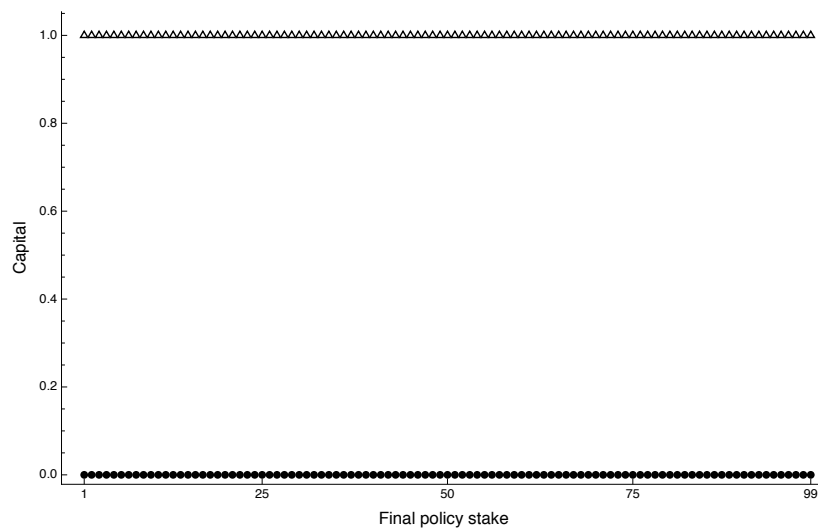

Figure 5: *Exercise 4.5*: Entire family of optimal policies for the Gambler's Problem for $p_h = 0.55$.