Problem 2.3

Define , for a given μ:

$$F_\mu(J)(i) = \frac{g(i,\mu(i)) + \alpha \sum_{j \neq i} p_{ij}(\mu(i))J(j)}{1 - \alpha p_{ii}(\mu(i))} = \frac{T_\mu(J)(i) - \alpha p_{ii}(\mu(i))J(i)}{1 - \alpha p_{ii}(\mu(i))}$$

Let $\epsilon > 0$. If $J \leq T(J)$, the we can find a μ such that $F_\mu(J) \leq F(J) + \epsilon e$. Then we have that:

$$F(J)(i) + \epsilon \geq F_\mu(J)(i) = \frac{T_\mu(J)(i) - \alpha p_{ii}(\mu(i))J(i)}{1 - \alpha p_{ii}(\mu(i))} \leq \frac{T_\mu(J)(i) + \epsilon - \alpha p_{ii}(\mu(i))J(i)}{1 - \alpha p_{ii}(\mu(i))} \leq T(J)(i) + \frac{\epsilon}{1-\alpha}$$

Thus $F(J)(i) \leq T(J)(i)$. Now since we have $J \leq T(J) \implies T(J) \leq F(J)$ and $J \geq T(J) \implies T(J) \geq F(J)$, we have that the fixed points of T and F are the same. Let J^* be the unique fixed point of F. Let $\delta > 0$. We have that

$$F(J + \delta e) \leq F(J) + \alpha \delta e$$

$$F(J) - \alpha \delta e \leq F(J - \delta e)$$

Since F is monotone, we have that $J \leq J' \implies F(J) \leq F(J')$. Now we pick a $\delta > 0$ such that $J - \delta e \leq J^* \leq J + \delta e$. By applying to F to each side k times, we obtain

$$F^k(J) - \alpha^k \delta e \leq J^* \leq F^k(J) + \alpha^k \delta e$$

Thus we have that $F^k(J) \to J^*$. Furthermore, $J \leq T(J) \implies T^k(J) \leq F^k(j) \leq J^*$ and $J \geq T(J) \implies T^k(J) \geq F^k(j) \geq J^*$, thus F converges faster than T.

For the second part of the question, note that setting $p_{ii} = 0$ gives:

$$(FJ)(i) = \min_{u \in U(i)} g(i,u) + \alpha \sum_j p_{ij}(u)J(j)$$

which is just the usual Bellman Equation for a dynamic programming problem with discount rate α.

Problem 3.2

a) The only improper policy is to never cheat. This means the game goes on forever with probability one. All other policies, which involve cheating at least some of the time, are proper, since the player is always caught with some positive probability. The cheating could require a certain outcome (tales is what makes sense) or a particular number of dollars

b) If we never cheat, we have the equations

$$V(H) = \frac{1}{2}(1 + V(H)) + \frac{1}{2}(0 + V(T))$$

$$V(T) = \frac{1}{2}(1 + V(H)) + \frac{1}{2}(-m + V(T))$$

Solving, we get $m = 2$. So for $m > 2$, the policy of never cheating has infinite negative value, and for $m < 2$, the game has infinite positive value if we never cheat. The expected value is 0 if $m = 2$.

c) First note that it is never optimal to cheat at H. Note that the value of cheating if H comes up, is:

$$V(H) = (1-p)(1+V(H))$$

Suppose we are at state H, we modify the policy so that we don't cheat at one stage, but do cheat at the next stage. Then we have:

$$\begin{aligned} V(H) &= \frac{1}{2}(1+V(H)) + \frac{1}{2}[(1-p)(1+V(H))] \\ &= \frac{1}{2} + \frac{1}{2}(V(H)+V(H)) \geq \frac{1}{2}V(H) \end{aligned}$$

So we see that wealth is improved by not cheating at stage H, and therefore we have that our policy must play fair at H. As we saw in part (b), never cheating when we have T and $m > 2$ gives negative infinite value. So we just have to calculate the value of always cheating. We have:

$$\begin{aligned} V(H) &= \frac{1}{2}[1+V(H)] + \frac{1}{2}V(T) \\ V(T) &= p(1-V(H)) \end{aligned}$$

Solving gives $V(H) = \frac{1}{p}$ and $V(T) = \frac{1-p}{p}$. Since these are both greater than negative infinity, we see that when $m > 2$, it is always optimal to cheat when tails and play fair when heads.

d) The same argument for never cheating when heads is still true. For cheating on tails, we have the same argument, only now note that $V(H)$ and $V(T)$ are less than positive infinity (which is the value of never cheating on tails), and therefore it is always optimal to play fair.

6.231 Dynamic Programming and Stochastic Control
Fall 2015