Puterman 3.26

This is a finite horizon MDP as we want to maximize the survival probability over T days. The lion can hunt once per day and thus the decision epochs are $t \in \{1, \ldots, T\}$.

The states should represent the lion's energy reserve. We therefore let the states represent the lion's gut fullness measured in kg's. Hence, we let $S' = [0, 30]$ and $\Delta = D$ represents death.

The actions correspond to selecting the size of the hunting group. Thus $A_s = \{0, 1, \ldots, M\}$, where M is the total number of lions in the pack. Of course, $A_\Delta = \{0\}$.

We are only interested in the survival of the lion. Therefore, we consider a binary reward function. Until time T, we have reward 0 as we cannot evaluate on the success until then. So, $r_t(s,a) = 0$ for $t < T$ and $\forall s, a$. The terminal value is given as $r_T(s) = 1$ if $s \in S'$ and $r_T(s) = 0$ if $s = \Delta$.

Finally, we need to specify the transition probabilities.

Puterman 3.26

We use the format

$\mathbb{P}(S_{t+1}=j \mid S_t=s, A_t=a)=p(j \mid s,a)$

For $s_t=\Delta$,

$p(\Delta \mid s_t, n)=1, \quad n \geq 0.$

For $0 \leq s_t < 6$,

$p(\Delta \mid s_t, 0)=1,$
$p(\Delta \mid s_t, n)=1-p(n), \quad n \geq 1,$
$p(\min(30, s_t-6-\frac{1}{2}+164/n) \mid s_t, n)=p(n), \quad n \geq 1.$

For $6 \leq s_t < 6\frac{1}{2}$

$p(s_t-6 \mid s_t, 0)=1,$
$p(\Delta \mid s_t, n)=1-p(n), \quad n \geq 1,$
$p(\min(30, s_t-6-\frac{1}{2}+164/n) \mid s_t, n)=p(n), \quad n \geq 1.$

For $6\frac{1}{2} \leq s_t \leq 30$,
$p(s_t-6 \mid s_t, 0)=1,$
$p(s_t-6-\frac{1}{2} \mid s_t, n)=1-p(n), \quad n \geq 1,$
$p(\min(30, s_t-6-\frac{1}{2}+164/n) \mid s_t, n)=p(n), \quad n \geq 1.$

Puterman 3.26

I'll add a few comments to these transition probabilities.

In the first case, we simply state that if the lion is dead today, then so will it be tomorrow.

In the second case, we state that if $0 \le s_t < 6$, then the lion will die if it does not hunt. Similarly, if it hunts unsuccessfully. If the lion has a succesfull hunt, it will use $1/2$ kg. on the hunt, 6 kg as its standard daily consumption and take in food; $164/n$ kg., and only stop if it is full, i.e. it has 30 kg. in the stomach. Given that it hunts in a group of n, the hunt is succesfull with probability $p(n)$.

This should be sufficient to understand the remaining cases.