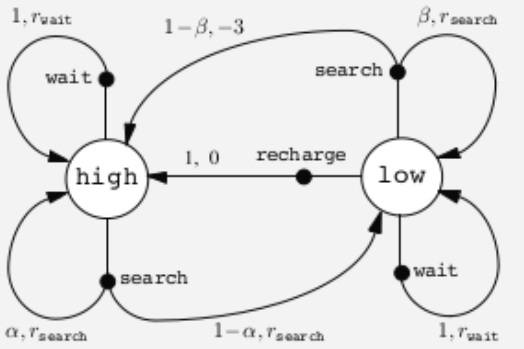
Ex. 3.4

s	a	s'	$p(s'\|s,a)$	$r(s,a,s')$
high	search	high	α	r_{search}
high	search	low	$1-\alpha$	r_{search}
low	search	high	$1-\beta$	-3
low	search	low	β	r_{search}
high	wait	high	1	r_{wait}
high	wait	low	0	-
low	wait	high	0	-
low	wait	low	1	r_{wait}
low	recharge	high	1	0
low	recharge	low	0	-



r_{search} and r_{wait} are average rewards. Immediate rewards actually come from the set

$R = \{-3, 1, 0\}$

battery depleted — can collected — no can collected

$$r(s,a,s') = \mathbb{E}\{R_{t+1} \mid S_t = s, A_t = a, S_{t+1} = s'\} = \sum_r r\, p(r|s,a,s') = \sum_r r \frac{p(s',r|s,a)}{p(s'|s,a)}$$

* $r(high, search, high) = 0 \frac{p(high,0|high,search)}{p(high|high,search)} + 1 \frac{p(high,1|high,search)}{p(high|high,search)} = \frac{p(high,1|high,search)}{\alpha} =$

$= r_{search} \Rightarrow \boxed{p(high,1|high,search) = \alpha\, r_{search}}$

* $r(high, search, low) = 0 \frac{p(low,0|high,search)}{p(low|high,search)} + 1 \frac{p(low,1|high,search)}{p(low|high,search)} = \frac{p(low,1|high,search)}{1-\alpha} =$

$= r_{search} \Rightarrow \boxed{p(low,1|high,search) = (1-\alpha)\, r_{search}}$

* $p(high|high,search) = p(high,0|high,search) + p(high,1|high,search) = p(high,0|high,search) + \alpha r_{search} =$

$= \alpha \Rightarrow \boxed{p(high,0|high,search) = \alpha(1 - r_{search})}$

$p(low|high,search) = p(low,0|high,search) + p(low,1|high,search) = p(low,0|high,search) + (1-\alpha) r_{search} =$

$= 1-\alpha \Rightarrow \boxed{p(low,0|high,search) = (1-\alpha)(1 - r_{search})}$

* $p(high \mid low, search) = \boxed{p(high, -3 \mid low, search) = 1 - \beta}$

* $r(low, search, low) = 0 \cdot \frac{p(low, 0 \mid low, search)}{p(low \mid low, search)} + 1 \cdot \frac{p(low, 1 \mid low, search)}{p(low \mid low, search)} = \frac{p(low, 1 \mid low, search)}{\beta} =$

$= r_{search} \Rightarrow \boxed{p(low, 1 \mid low, search) = \beta\, r_{search}}$

* $p(low, 0 \mid low, search) + p(low, 1 \mid low, search) = p(low \mid low, search)$

$p(low, 0 \mid low, search) + \beta\, r_{search} = \beta$

$\boxed{p(low, 0 \mid low, search) = (1 - r_{search})\, \beta}$

* $r(high, wait, high) = 0 \cdot \frac{p(high, 0 \mid high, wait)}{p(high \mid high, wait)} + 1 \cdot \frac{p(high, 1 \mid high, wait)}{p(high \mid high, wait)} = \frac{p(high, 1 \mid high, wait)}{1} = r_{wait} \Rightarrow$

$\Rightarrow \boxed{p(high, 1 \mid high, wait) = r_{wait}}$

* $p(high, 0 \mid high, wait) + p(high, 1 \mid high, wait) = p(high \mid high, wait)$

$p(high, 0 \mid high, wait) + r_{wait} = 1$

$\boxed{p(high, 0 \mid high, wait) = 1 - r_{wait}}$

* $r(low, wait, low) = 0 \cdot \frac{p(low, 0 \mid low, wait)}{p(low \mid low, wait)} + 1 \cdot \frac{p(low, 1 \mid low, wait)}{p(low \mid low, wait)} = \frac{p(low, 1 \mid low, wait)}{1} = r_{wait} \Rightarrow$

$\Rightarrow \boxed{p(low, 1 \mid low, wait) = r_{wait}}$

* $\boxed{p(low, 0 \mid low, wait) = 1 - r_{wait}}$ (Idem to $p(high, 0 \mid high, wait)$)

* $\boxed{p(high, 0 \mid low, recharge) = 1}$

Final table

s'	r	s	a	$p(s', r \mid s, a)$
high	1	high	search	αr_{search}
high	0	high	search	$\alpha (1 - r_{search})$
low	1	high	search	$(1-\alpha) r_{search}$
low	0	high	search	$(1-\alpha)(1-r_{search})$
high	-3	low	search	$1-\beta$
low	1	low	search	βr_{search}
low	0	low	search	$\beta (1-r_{search})$
high	0	low	recharge	1