Assignment 6
Theoretical Neuroscience

Peter Dayan

November 28, 2008

Please note that the equation numbers refer to equations in Dayan and Abbott, chapter 9.

1. Consider the case of partial reinforcement (studied in figure 9.1) in which reward $r = 1$ is provided randomly with probability p on any given trial. Assume that there is a single stimulus with $u = 1$, so that $\epsilon\delta u$, with $\delta = r - v = r - wu$, is equal to $\epsilon(r - w)$. By considering the expected value $\langle w + \epsilon(r - w)\rangle$ and the expected square value $\langle (w + \epsilon(r - w))^2\rangle$ of the new weights, calculate the self-consistent equilibrium values of the mean and variance of the weight w. What happens to your expression for the variance if $\epsilon = 2$ or $\epsilon > 2$? To what features of the learning rule do these effects correspond?

2. Implement a stochastic three-armed bandit using the indirect actor and the action choice softmax rule 9.12. Let arm a produce a reward of p_a, with $p_1 = 1/4, p_2 = 1/2, p_3 = 3/4$, and use a learning rate of $\epsilon = 0.01, 0.1, 0.5$ and $\beta = 1, 10, 100$. Consider what happens if after every 250 trials, the arms swap their reward probabilities at random. Averaging over a long run, explore to see which values of ϵ and β lead to the greatest cumulative reward. Can you account for this behavior?

3. Repeat the above excercise using the direct actor (with learning rule 9.22). For $\bar{r}$, use a low-pass filtered version of the actual reward, which is obtained by using the update rule
$$\bar{r} \rightarrow \lambda\bar{r} + (1 - \lambda)r$$
with $\lambda = 0.95$. Study the effect of the different values of ϵ and β in controlling the average rate of rewards when the arms swap their reward probabilities at random every 250 trials.