Homework: $n-$armed bandit (Week 7)

1. (An in-class exercise) What does this code snippet do?

```
p=[0.1,0.4,0.5];
BinEdges=[0 cumsum(p(:))']

x=rand(15,1);
[A,B]=histcounts(x,BinEdges)
```

   Problem: Notice that we need to index every value in x as to which bin it is in. That's why we need to loop it instead. Try this:

```
x=rand(15,1);

for j=1:15
 A=histcounts(x(j),BinEdges)
end
```

2. (This is the exercise on p 26 of your notes) Suppose we have two returns, Q_1 and Q_2 (with $Q_1 > Q_2$). Suppose we take P_1 and P_2 to be from the softmax algorithm,

$$P_1 = \frac{\exp(Q_1/\tau)}{\exp(Q_1/\tau) + \exp(Q_2/\tau)}$$

$$P_2 = \frac{\exp(Q_2/\tau)}{\exp(Q_1/\tau) + \exp(Q_2/\tau)}$$

   Take the limit as $\tau \to 0$ and $\tau \to \infty$. (We'll do one together in class, for homework, do them both.

3. (This is from the top of p. 28 of your notes) Consider the two methods in class of decreasing some value from a to b in N steps:

$$f_1(t) = a + \frac{b-a}{N}t$$

$$f_2(t) = a\left(\frac{b}{a}\right)^{t/N}$$

   If $a = 10$ and $b = 1$, and $N = 15$ get a plot of the two functions. Think about what your domain ought to be, and explain why one of these might work better than the other.

4. (These are from the bottom of p. 28 of your notes, with better wording) Suppose we have three estimated returns, Q_1, Q_2, Q_3 and Q_1 is the best of the three. Show that, given associated values of P_1, P_2, P_3, the updated values will still all sum to 1.

5. (From p. 29) Suppose that for some fixed machine a, the return $Q_t(a)$ is never the maximum. Show that, by using the update rule, the corresponding probability $P_t(a)$ goes to zero as $t \to \infty$. HINT: Show that

$$P_t(a) = (1 - \beta)^t P_0(a)$$

where $P_0(a)$ is the initial probability.