A Statistical Link between Learning and Evolution

Brendan Kitts *
Center for Complex Systems,
Brandeis University, Waltham, MA 02254.
brendy@cs.brandeis.edu

Abstract

Widely replicated experimental findings show that faced with multiple variable reward outcomes, animals sample each proportional to their mean payoff. This finding is explained computationally using the Holland theorem, in which this style of sampling is optimal given certain assumptions about the distribution. Other adaptive phenomena which seem to be consistent with this law are discussed, and it is suggested that the 'adaptive problem' may be broadly similar across different domains.

1 Response Matching: A curious behavioural constant

The *matching law* was first uncovered by Richard Herrnstein in operant conditioning experiments in the 1950s. Herrnstein was examining how pigeons behave on a concurrent reinforcement schedule pecking at two response keys, k_1 and k_2. Reinforcement was supplied at a different rate for each key, given by $r_1(k_1)$ and $r_2(k_2)$. Herrnstein found that the relative frequency of responding, b_1 to k_1 and b_2 to k_2 was equal to the relation:

$$\frac{b_1}{b_2} \propto \frac{r_1}{r_2} \tag{1}$$

*supported by ONR grant no. N00014-95-1-0759

Soon after this result research into the area exploded, with the same finding replicated with only small deviations across a number of labs, preparations and species (for a good survey see Baum, 1979). For many years however, the finding remained puzzling, as given two rewards it would seem that a higher overall reward could be achieved just by sampling the higher-payoff alternative. Why do animals continue to sample from *both* at these specific frequencies? The answer may be found in a sampling theorem derived from the genetic algorithms literature some 25 years later.

2 The Holland Theorem

The problem of maximising payoff from two or more stochastic processes is a classic problem, which may be formalised as a *K-armed bandit* problem. Imagine a person walks into a room with K slot machines, each of which may be biased differentially. The problem is to maximise this person's expected payoff from these machines, from an outlay of only N trials (or coins).

One strategy might be to simply allocate all the trials to the machine with the highest observed payoff, say after t trials. Unfortunately, if after this time, the machine's payoff was the result of some "lucky pulls", then by settling on this machine our person can expect to incur a loss over the following trials equal to $N(u_1 - u_2)$, ie. the difference between the mean of the true best and selected choice, multiplied by every trial the erroneous sampling is performed.

Holland (1975) investigated this scenario in the context of trying to develop a search strategy for genetic algorithms (where different individuals had various payoffs, and the algorithm had to decide which individuals should be retained as good solutions). The theoretical result he derived would show that payoff could be maximised by sampling at an exponentially higher rate between an observed higher and lower payoff alternative, and this can be approximated by sampling proportional to fitness.

3 Optimal allocation of trials

Assume we have N trials to allocate between two random variables. Let $K_{(1)}$ be the observed highest payoff, and $K_{(2)}$ be the observed second highest, with observed means $u_1 > u_2$ and variances σ_1^2 and σ_2^2. We now want to choose some value of n trials to allocate to $K_{(2)}$ and $N - n$ trials to $K_{(1)}$ so as to minimise future expected loss.

3.1 Theorem

Expected loss is minimised when $N - n = \sqrt{8\pi} \cdot b \cdot e^{\left[\frac{b^{-2}n + \lg n}{2}\right]}$, which is approximated by setting $N - n \propto$ observed payoff $K_{(1)}$.

3.2 Proof

The loss equation which expresses how much we loose from dividing $N - n$ and n trials between the two random variables, is equal to

$$L(N-n,n) = [q(N-n,n)(N-n) + (1 - q(N-n,n))n]|u_1 - u_2| \quad (2)$$

where q is the probability that the observed best so far is actually second best. This follows from the standard formulation in decision theory. The loss equation states that our expected loss will be equal to (trials expended on best * probability that its second best) + (trials expended on second best * probability that its the best), multiplied by the difference between the means of the best and second best. Our task is to select n which minimises L. We will do this as follows: First we will determine q, which involves calculating it using Bayes rule and getting its value over repeated samples using the Central Limit Theorem. We will then put this value into our loss equation, L, set $\frac{dL}{dn} = 0$, and solve to find the value of n which minimises L.

3.2.1 derive Bayes expression for $q(N-n,n)$

We first break $q(N-n,n)$ into a Bayes probability density, which involves introducing the following quantities:

$q' = Pr(K' = K_{(2)}|K' = K_1)$ This is the probability that K' is observed to be the second highest, given that its actually the best. This probability can also be restated as $q' = Pr(\frac{1}{n}\sum_1^n K_1 < \frac{1}{N-n}\sum_1^{N-n} K_2)$ which means that the mean payoffs for K_1 are less than the mean payoffs for K_2.

$q'' = Pr(K' = K_{(2)}|K' = K_2)$ This is the probability that K' is observed to be the second highest, given that it is second highest.

$p = Pr(K' = K_1)$ The prior probability (*a priori* bias) that the variable K' is the best. Without additional information we will eventually just write this as 0.5.

$1 - p = Pr(K' = K_2)$ The prior probability that K' is second best. The value for $1 - p = 0.5$.

By Bayes theorem, we write the probability for q as

$$q(N-n,n) = \frac{q'p}{q'p + q''(1-p)} \quad (3)$$

3.2.2 use the Central Limit Theorem to show that over series of trials, q' approaches normal

By the central limit theorem, $\frac{1}{N-n}\sum_1^{N-n} K_2$ (payoff from actual second highest) approaches a normal distribution with mean u_2 and variance $\frac{\sigma_2^2}{N-n}$. $\frac{1}{n}\sum_1^n K_1$ (payoff from actual highest) approaches normal with mean u_1 and variance $\frac{\sigma_1^2}{n}$. The difference between these two normal distributions (giving us the probability that the mean payoffs from $K_1 < K_2$ or $Pr(\frac{1}{n}\sum_1^n K_1 < \frac{1}{N-n}\sum_1^{N-n} K_2)$, is equal to a normal with mean $u_1 - u_2$ and variance $\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{N-n}$. Thus q' or $Pr(\frac{1}{n}\sum_1^n K_1 - \frac{1}{N-n}\sum_1^{N-n} K_2 < 0)$ the probability that $K_{(1)}$ achieves an average payoff less than $K_{(2)}$ is equal to $N[(u_1 - u_2), (\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{N-n})]$, or when normalised to mean 0 and unit variance (z-score),

$$q' = N[\frac{u_1 - u_2}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{N-n}}}] \tag{4}$$

3.2.3 introduce equation for normal and drop small terms

The tail of a normal is equal to $\frac{1}{\sqrt{2\pi}} \cdot \frac{e^{\frac{-x^2}{2}}}{x}$. Therefore, substituting in the parameters of our normal distribution for q' (**??**) we derive the expansion $q' = \frac{1}{\sqrt{2\pi}} \cdot \frac{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{N-n}}}{u_1-u_2} \cdot e^{\frac{1}{2}\frac{-(u_1-u_2)^2}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{N-n}}}}$. Analogous to the above derivation of q', we can also derive q'', obtaining: $q'' = \frac{1}{\sqrt{2\pi}} \cdot \frac{\sqrt{\frac{\sigma_1^2}{N-n}+\frac{\sigma_2^2}{n}}}{u_1-u_2} \cdot e^{\frac{1}{2}\frac{-(u_1-u_2)^2}{\sqrt{\frac{\sigma_1^2}{N-n}+\frac{\sigma_2^2}{n}}}}$.

By inspection, as n increases, both q' and $1 - q''$ decrease exponentially, becomming very small. Therefore we can approximate the Bayes expression for q (**??**) as $\frac{q'p}{q'p+q''(1-p)} \approx q'(\frac{p}{1-p})$. With the assumption that $p = 0.5$ (our prior knowledge about each), we get

$$q(N-n, n) \approx q' \tag{5}$$

3.2.4 Find n at $\frac{dL}{dt} = 0$

We want to know how the expected loss changes with respect to an increase or decrease in n. By setting this to zero, we can find the value of n which maximises L.

$\frac{dL}{dn} = |u_1-u_2|[-q+(N-n)\frac{d}{dn}q+1-q-n\frac{dq}{dn}] = |u_1-u_2|[(1-2q)+(N-2n)\frac{dq}{dn}]$
$= |u_1 - u_2|[(1-2q) + (N-2n)q\frac{x^2+1}{2n}]$

when $\frac{dL}{dt} = 0$ we get $0 = (1-2q) - \frac{N-2n}{2n}q(x^2+1)$, or $\frac{N-2n}{2n} = \frac{1-2q}{q(x^2+1)}$. Expanding using our value for q, (**??**) ($q \approx q'$) this becomes $\frac{N-2n}{n} = \frac{2\sigma_1\sqrt{2\pi}}{u_1-u_2} \cdot \frac{1}{\sqrt{n}} \cdot e^{\frac{(u_1-u_2)^2 n}{2\sigma_1^2}}$. Introducing $b = \sigma_1(u_1 - u_2)$ for simplification, we obtain our solution $N - n = \sqrt{8\pi} \cdot b \cdot e^{\frac{b^{-2}n + \lg n}{2}}$. This verifies the first part of the theorem. It can be observed that since the cost of the RHS (trials allocated to the observed poorer alternative) grows exponentially faster than LHS, then $N - n$, the trials allocated to the better observed variable, should be increased exponentially compared to n in order to maintain the equality. Without worrying too much about the precise order of growth, this will be the case when $\frac{d}{dt}N - n = cK_{(1)}$ since $K_{(1)} > K_{(2)}$. Thus, a sampling regime in which $N - n$ is sampled proportional to the observed payoff of $K_{(1)}$ will approximately minimise loss.

4 Biological applications of the theorem

The Hebbian neural learning rule (or "outstar") can provide a neurally plausible implementation of the matching algorithm. Consider elements of a set of behaviours, $b \in B$ each with an associated probability of being selected, $P(b)$. After sampling, a reward $R(b)$ is calculated. Probability of selection can be increased as follows (Grossberg, 1975; Levy and Desmond, 1985):

$$\frac{dP}{dt} = [R(b) - P(b)]\alpha \tag{6}$$

where α is an arbitrary constant $1 \geq \alpha \geq 0$. In the limit, the probabilities $P(b)$ of response approach the true reward of the environment $R(b)$. At any time therefore, sampling of behaviour is proportional to reward.

The theorem also crops up in some intriguing biological areas. According to modern evolutionary theory, evolution works by organisms reproducing proportional to their fitness. If this is true, then there may be an important similarity between ethology and evolution: both use the same sampling method.

5 Conclusion

Its heartening that a significant behaviourist finding from the 1950s, which spurred research and then lay almost forgotten in the intervening paradigm

shifts, could 40 years later be explained in terms of sampling optimality. Indeed, the only mention of Herrnstein's (widely replicated) data today that the author has been aware of, was in passing as an example of the brain's "puzzling" disposition towards contiguity in a modern neuroscience text (Thompson, 1993; pp. 343). The contempory work of John Holland and others in the neural networks and statistics communities provides an excellent example of the importance of basic mathematical research, and its potential to cross-fertilize with other disciplines and lead to "deep understandings" of seemingly disparate processes.

References

[1] Baum, W. (1979), "Matching, undermatching, and overmatching in studies of choice", *Journal of the Experimental Analysis of Behaviour*, Vol. 32, No. 2, pp. 269-281.

[2] Grossberg, S. (1976), "On the Development of Feature Detectors in the Visual Cortex with Applications to Learning and Reaction-Diffusion Systems", *Biological Cybernetics*, Vol. 21, pp. 145-159.

[3] Herrnstein, R. (1958), "Some factors influencing behaviour in a two-response situation", *Transactions of the New York Academy of Science*, Vol. 21, pp. 35-45.

[4] Holland, J. (1992), *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann-Arbor.

[5] Levy, W. and Desmond, N. (1985), "The Rules of Elemental Synaptic Plasticity", from Levy, W., Anderson, J. and Lehmkuhle, S. (eds), *Synaptic Modification, Neuron Selectivity, and Nervous System Organisation*, Lawrence Erlbaum Associates.

[6] Thompson, R. (1993), *The Brain*, W.H.Freeman and Co., New York.