Decision theory – a simple example

David MacKay, Cavendish Laboratory, August 6, 2001

Introduction

Decision theory is trivial, apart from computational details.

You have a choice of various actions, a. The world may be in one of many states $\mathbf{x}$; which one occurs may be influenced by your action. The world's state is described by a probability distribution $P(\mathbf{x}|a)$. Finally, there is a utility function $U(\mathbf{x}, a)$ which specifies the payoff you receive when the world is in state $\mathbf{x}$ and you chose action a.

The task of decision theory is to select the action that maximizes the expected utility,

$$\mathcal{E}[U|a] = \int d^K\mathbf{x}\, U(\mathbf{x}, a) P(\mathbf{x}|a). \tag{1}$$

That's all. The computational problem is to maximize $\mathcal{E}[U|a]$ over a. [Pessimists may prefer to define a loss function L instead of a utility function U and minimize the expected loss.]

Is there anything more to be said about decision theory?

Well, in a real problem, the choice of an appropriate utility function may be quite difficult. Furthermore, when a sequence of actions is to be taken, with each action providing information about $\mathbf{x}$, we have to take into account the affect that this anticipated information may have on our subsequent actions. The resulting mixture of forward probability and inverse probability computations in a decision problem is distinctive. In a realistic problem such as playing a board game, the tree of possible cogitations and actions that must be considered becomes enormous, and 'doing the right thing' is not simple (Russell and Wefald, 1991; Baum and Smith, 1993; Baum and Smith, 1997) because the expected utility of an action cannot be computed exactly.

Perhaps a simple example is worth exploring.

Rational prospecting

Suppose you have the task of choosing the site for a Tanzanite mine. Your final action will be to select the site from a list of N sites. The nth site has a net value called the return x_n which is initially unknown, and will be found out exactly only after site n has been chosen. [x_n equals the revenue earned from selling the Tanzanite from that site, minus the costs of buying the site, paying the staff, and so forth.] At the outset, the return x_n has a probability distribution $P(x_n)$, based on the information already available.

Tanzanite is a mineral found in East Africa.

Before you take your final action you have the opportunity to do some prospecting. Prospecting at the nth site has a cost c_n and yields data d_n which reduce the uncertainty about x_n. [We'll assume that the returns of the N sites are unrelated to each other, and that prospecting at one site only yields information about that site and doesn't affect the return from that site.]

Your decision problem is:

given the initial probability distributions $P(x_1)$, $P(x_2)$, $\ldots P(x_N)$, first, decide whether to prospect, and at which sites; then choose which site to mine.

For simplicity, let's make everything in the problem Gaussian and focus on the question of whether to prospect once or not. We'll assume our utility function is linear in x_n; we wish to maximize our expected return. The utility function is

$$U = x_{n_a}, \tag{2}$$

if no prospecting is done, where n_a is the chosen 'action' site, and if prospecting is done the utility is

$$U = -c_{n_p} + x_{n_a}, \tag{3}$$

where n_p is the site at which prospecting took place.

The notation $P(y) = \text{Normal}(y; \mu, \sigma^2)$ indicates that y has Gaussian distribution with mean μ and variance σ^2.

The prior distribution of the return of site n is

$$P(x_n) = \text{Normal}(x_n; \mu_n, \sigma_n^2). \tag{4}$$

If you prospect at site n, the datum d_n is a noisy version of x_n:

$$P(d_n|x_n) = \text{Normal}(d_n; x_n, \sigma^2). \tag{5}$$

Exercise 1: Given these assumptions, show that the prior probability distribution of d_n is

$$P(d_n) = \text{Normal}(d_n; \mu_n, \sigma^2 + \sigma_n^2) \tag{6}$$

[mnemonic: when independent variables add, variances add], and that the posterior distribution of x_n given d_n is

$$P(x_n|d_n) = \text{Normal}\left(x_n; \mu_n', {\sigma_n^2}'\right) \tag{7}$$

where

$$\mu_n' = \frac{d_n/\sigma^2 + \mu_n/\sigma_n^2}{1/\sigma^2 + 1/\sigma_n^2} \quad \text{and} \quad \frac{1}{{\sigma_n^2}'} = \frac{1}{\sigma^2} + \frac{1}{\sigma_n^2} \tag{8}$$

[mnemonic: when Gaussians multiply, precisions add].

To start with let's evaluate the expected utility if we do no prospecting (*i.e.*, choose the site immediately); then we'll evaluate the expected utility if we first prospect at one site and then make our choice. From these two results we will be able to decide whether to prospect once or zero times, and, if we prospect once, at which site.

So, first we consider the expected utility without any prospecting.

Exercise 2: Show that the optimal action, assuming no prospecting, is to select the site with biggest mean

$$n_a = \underset{n}{\text{argmax}}\ \mu_n, \tag{9}$$

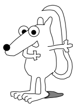
and the expected utility of this action is

$$\mathcal{E}[U|\text{optimal } n] = \max_n \mu_n. \tag{10}$$

[If your intuition says 'surely the optimal decision should take into account the different uncertainties σ_n too?', the answer to this question is 'reasonable – if so, then the utility function should be *nonlinear* in x'.]

Now the exciting bit. Should we prospect? Once we have prospected at site n_p, we will choose the site using the descision rule (9) with the value of mean μ_{n_p} replaced by the updated value μ'_n given by (8). What makes the problem exciting is that we don't yet know the value of d_n, so we don't know what our action n_a will be; indeed the whole value of doing the prospecting comes from the fact that the outcome d_n may alter the action from the one that we would have taken in the abscence of the experimental information.

From the expression for the new mean in terms of d_n (8), and the known variance of d_n (6) we can compute the probability distribution of the key quantity, μ'_n, and can work out the expected utility by integrating over all possible outcomes and their associated actions.



Exercise 3: Show that the probability distribution of the new mean μ'_n (8) is Gaussian with mean μ_n and variance

$$s^2 \equiv \sigma_n^2 \frac{\sigma_n^2}{\sigma^2 + \sigma_n^2}. \tag{11}$$

Consider prospecting at site n. Let the biggest mean of the other sites be μ_1. When we obtain the new value of the mean, μ'_n, we will choose site n and get an expected return of μ'_n if $\mu'_n > \mu_1$, and we will choose site 1 and get an expected return of μ_1 if $\mu'_n < \mu_1$.

So the expected utility of prospecting at site n, then picking the best site, is

$$\mathcal{E}[U|\text{prospect at } n] = -c_n + P(\mu'_n < \mu_1)\, \mu_1 + \int_{\mu_1}^{\infty} d\mu'_n\, \mu'_n \,\text{Normal}(\mu'_n; \mu_n, s^2). \tag{12}$$

The difference in utility between prospecting and not prospecting is a quantity of interest, and it depends on what we would have done without prospecting. If μ_1 is not only the biggest of the rest, but is also bigger than μ_n, then we would have chosen μ_1; if μ_n, we would have chosen n.

$$\mathcal{E}[U|\text{no prospecting}] \;=\; \begin{cases} -\mu_1 & \text{if } \mu_1 \geq \mu_n \\ -\mu_n & \text{if } \mu_1 \leq \mu_n \end{cases} \tag{13}$$

So

$$\begin{aligned} &\mathcal{E}[U|\text{prospect at } n] - \mathcal{E}[U|\text{no prospecting}] \\ &= \begin{cases} -c_n + \displaystyle\int_{\mu_1}^{\infty} d\mu'_n\, (\mu'_n - \mu_1)\,\text{Normal}(\mu'_n; \mu_n, s^2) & \text{if } \mu_1 \geq \mu_n \\ -c_n + \displaystyle\int_{-\infty}^{\mu_1} d\mu'_n\, (\mu_1 - \mu'_n)\,\text{Normal}(\mu'_n; \mu_n, s^2) & \text{if } \mu_1 \leq \mu_n. \end{cases} \end{aligned} \tag{14}$$

We can plot the change in expected utility due to prospecting (omitting c_n) as a function of (horizontal axis) the difference $(\mu_n - \mu_1)$ and (vertical axis) the initial standard deviation σ_n. In the figure the noise variance is $\sigma^2 = 1$.

References

Baum, E. B., and Smith, W. D. (1993) Best play for imperfect players and game tree search. Technical report, Princeton, NJ.

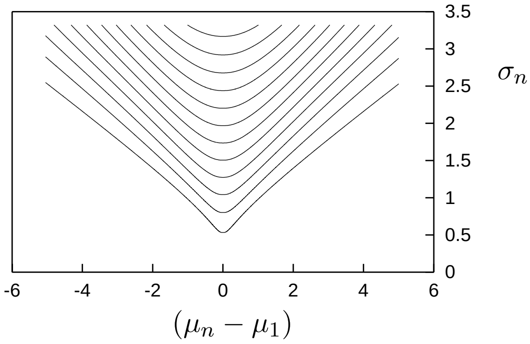


Figure 1: The gain in expected utility due to prospecting. The contours are equally spaced from 0.1 to 1.2 in steps of 0.1. To decide whether it is worth prospecting at site n, find the contour equal to c_n; all points $[(\mu_n - \mu_1), \sigma_n]$ above that contour are worthwhile.

Baum, E. B., and Smith, W. D. (1997) A Bayesian approach to relevance in game playing. *Artificial Intelligence* **97** (1-2): 195–242.

Russell, S., and Wefald, E. (1991) *Do the Right Thing: Studies in Limited Rationality*. MIT Press.