CHAPTER 4

VALUE OF INFORMATION POLICIES

An important class of problems are those where experiments are expensive, and as a result the budget for the number of experiments is relatively small. For now, we are going to stay with the same types of problems we addressed in chapter 3, which is to say that we have a not-too-large set of discrete alternatives, described by a lookup table belief model. However, now we are going to assume that are experiments are expensive, which translates to relatively small budgets, which may be smaller (sometimes much smaller) than the number of alternatives.

At the same time, we are going to relax what are sometimes strict conditions on how much time we have to determine what to do next. In internet applications, we may have 50 milliseconds to determine what to do next. In the context of expensive decisions, we are not going to object to decision rules that might take seconds or minutes (or more) to compute.

Examples of expensive experiments include

- We may drill a well to evaluate the potential of the underground geology for producing oil or gas.
- We may need to observe a patient taking a new drug for several weeks to see how her body responds to the medication.
- A business may need to observe the sales of a product over a period of several weeks to evaluate the market response to a price.

Optimal Learning. By Warren B. Powell and Ilya O. Ryzhov

- A basketball coach needs to observe how well a team of five players performs over a course of several games.
- A scientist may require a day (or longer) to run a single experiment to assess the impact of a particular experimental design on the strength or conductivity of a material.

In the realm of high volume information (as is often the case in internet applications), we can often run a series of exploratory tests. When we face the problem of expensive experiments, then we find ourselves thinking carefully about the first experiment, before we have collected any information. However, it is often the case that we already know something about the properties of the problem.

For this reason, we are going to adopt a Bayesian modeling approach so that we have a mechanism for incorporating prior knowledge, which tends to almost always be available for these complex problems. Priors can come from a variety of sources:

Prior experiments We may have already run experiments, perhaps in other settings.

Research literature Others may have worked on this or related problems.

Domain expertise Experts may offer initial insights, often drawn from experiences with similar settings (such as the response of other patients to a drug, or the market response to the price of a textbook).

Computer simulations We may learn from numerical models.

A valuable class of policies for expensive experiments is based on maximizing the value of information. In chapter 3, we optimized policies to maximize $\overline{F}^\pi$, but this simply means doing the best we can with a particular heuristic. In this chapter, we are going to design policies that are specifically designed to maximize the performance from *each* experiment so that we obtain the fastest learning rate possible.

4.1 THE VALUE OF INFORMATION

There are different ways of estimating the value of information, but one strategy, called the *knowledge gradient*, works as follows. Assume that we have a finite number of discrete alternatives with independent, normally distributed beliefs (the same problem we considered in chapter 3). After n experiments, we let $\bar{\mu}^n$ be our vector of estimates of means and β^n our vector of precisions (inverse variances). We represent our belief state as $S^n = (\bar{\mu}^n_x, \beta^n_x)_{x\in\mathcal{X}}$. If we stop measuring now, we would pick the best option, which we represent by

$$x^n = \arg\max_{x\in\mathcal{X}} \bar{\mu}^n_x.$$

The value of being in state S^n is then given by

$$V^n(S^n) = \bar{\mu}^n_{x^n}. \tag{4.1}$$

Now let $S^{n+1}(x)$ be the next state if we choose to measure $x^n = x$ right now, allowing us to observe $W^{n+1}_{x^n}$. This allows us to update $\bar{\mu}^n_x$ and β^n_x, giving us an estimate $\bar{\mu}^{n+1}_x$ for the mean and β^{n+1}_x for the precision (using equations (3.2) and (3.3)). Given that we choose to measure $x = x^n$, we transition to a new belief state $S^{n+1}(x)$, and the value of being in this state is now given by

$$V^{n+1}(S^{n+1}(x)) = \max_{x' \in \mathcal{X}} \bar{\mu}^{n+1}_{x'}(x).$$

Let $\bar{\mu}^{n+1}_{x'}(x)$ be the updated estimate of $\bar{\mu}^n_{x'}$ if we run experiment x' and observe $W^{n+1}_{x'}$.

$$\bar{\mu}^{n+1}_{x'}(x) \quad = \quad \begin{cases} \frac{\beta^n_{x'}\bar{\mu}^n_{x'} + \beta^W W^{n+1}_{x'}}{\beta^n_{x'} + \beta^W} & \text{If } x' = x, \\ \bar{\mu}^n_{x'} & \text{Otherwise} \end{cases} . \tag{4.2}$$

At time n when we have chosen $x = x^n$ as our next experiment, but before we have observed W^{n+1}_x, the estimate $\bar{\mu}^{n+1}_{x'}(x)$ is a random variable.

We would like to choose x at iteration n which maximizes the expected value of $V^{n+1}(S^{n+1}(x))$. At time n, we write this expectation as

$$\mathbb{E}\{V^{n+1}(S^{n+1}(x))|S^n\} \quad = \quad \mathbb{E}_\mu \mathbb{E}_{W|\mu}\{V^{n+1}(S^{n+1}(x))|S^n\},$$

where the right hand side brings out that there are two random variables: the truth μ (which is uncertain in a Bayesian belief model) and the outcome $W^{n+1}_x = \mu_x + \epsilon^{n+1}$, which depends on the unknown truth μ. We want to choose an experiment x that maximizes $\mathbb{E}\{V^{n+1}(S^{n+1}(x))|S^n\}$, but instead of writing our objective in terms of maximizing the value from an experiment, we are going to write it equivalent as maximizing the incremental value from the experiment, which is given by

$$\begin{aligned} \nu^{KG,n}_x &= \mathbb{E}_\mu \mathbb{E}_{W|\mu}\left\{V^{n+1}(S^{n+1}(x)) - V^n(S^n)|S^n\right\} \\ &= \mathbb{E}_\mu \mathbb{E}_{W|\mu}\left\{V^{n+1}(S^{n+1}(x))|S^n\right\} - V^n(S^n). \end{aligned} \tag{4.3}$$

Keep in mind that the state S^n is our belief about μ after n experiments, which is that $\mu \sim N(\bar{\mu}^n, \beta^n)$. Given S^n (that is, given what we know at time n), the value $V^n(S^n)$ is calculated just using our current estimates $\bar{\mu}^n$, as is done in equation (4.1). Thus, at time n, $V^n(S^n)$ is a number, which is why $\mathbb{E}\{V^n(S^n)|S^n\} = V^n(S^n)$. However, $V^{n+1}(S^{n+1}(x))$ is a random variable since it depends on the outcome of the $n+1st$ experiment W^{n+1}.

The right hand side of (4.3) can be viewed as the derivative (or gradient) of $V^n(S^n)$ with respect to the experiment x. Thus, we are choosing our experiment to maximize the gradient with respect to the knowledge gained from the experiment. For this reason we refer to $\nu^{KG,n}_x$ as the *knowledge gradient*. We write the knowledge gradient policy using

$$X^{KG,n} = \arg\max_{x \in \mathcal{X}} \nu^{KG,n}_x. \tag{4.4}$$

The knowledge gradient, $\nu^{KG,n}$, is the amount by which the solution improves if we choose to measure alternative x. This is illustrated in Figure 4.1, where the estimated mean of choice 4 is best, and we need to find the value from measuring choice 5. The

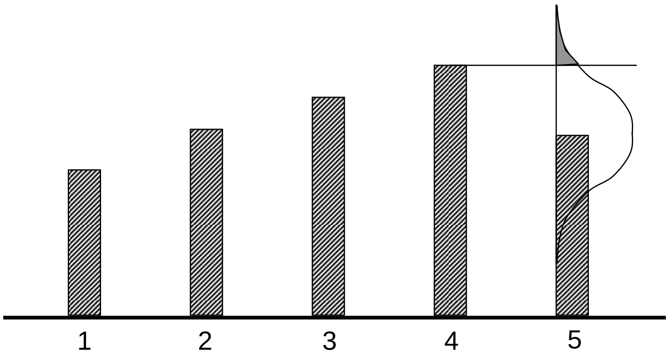


Figure 4.1 Illustration of the knowledge gradient if we were to measure choice 5.

estimated mean of choice 5 will move up or down according to a normal distribution (we assume with mean 0). The solid area under the curve that exceeds the estimate for choice 4 is the probability that measuring 5 will produce a value that is better than the current best, which means that V^{n+1} will increase. The knowledge gradient is the expected amount by which it will increase (we receive a value of 0 if it does not go up).

4.2 THE KNOWLEDGE GRADIENT FOR INDEPENDENT BELIEFS

For the case of independent normally distributed beliefs, the knowledge gradient is particularly easy to compute. When independence holds, we only change our beliefs about one alternative in every time period. Suppose that we are at time n, with estimates $\bar{\mu}_x^n$ of the true values μ_x, after which we choose to measure a particular alternative x^n. Then, we will have $\bar{\mu}_x^{n+1} = \bar{\mu}_x^n$ for $x \neq x^n$. Furthermore, for $x = x^n$, while the exact value of $\bar{\mu}_x^{n+1}$ is still unknown at time n, there is a very simple expression for the *conditional* distribution of $\bar{\mu}_x^{n+1}$ given the time-n beliefs. We are able to write

$$\bar{\mu}_x^{n+1} \sim \mathcal{N}\left(\bar{\mu}_x^n, \tilde{\sigma}_x^{2,n}\right) \tag{4.5}$$

where $\tilde{\sigma}_x^{2,n}$ is conditional change in the variance (which we first saw in chapter 2) given by

$$\begin{aligned}\tilde{\sigma}_x^{2,n} &= \mathit{Var}^n[\bar{\mu}_x^{n+1}] \\ &= \mathit{Var}^n[\bar{\mu}_x^{n+1} - \bar{\mu}_x^n].\end{aligned} \tag{4.6}$$

We can think of $\tilde{\sigma}_x^{2,n}$ as the variance of the change in the estimate, or the conditional variance of $\bar{\mu}_x^{n+1}$ (given what we know at time n).

The distribution in (4.5) is known as the *predictive distribution* of $\bar{\mu}_x^{n+1}$, because it represents our best prediction of the results of our next observation before the observation actually occurs.

Below we provide the calculations required to compute the knowledge gradient, and follow this presentation with a discussion of some properties of this policy. The full derivation of the knowledge gradient policy is deferred to Section 4.12.1. For now, it is enough to keep in mind that (4.3) involves computing an expected value over the predictive distribution.

4.2.1 Computation

For the case of independent normally distributed beliefs, the knowledge gradient is particularly easy to compute. Recall that the precision is simply the inverse of the variance, which is given by $\bar{\sigma}_x^{2,n}$. Making the transition from precisions to variances, let σ_W^2 be the variance of our experiment W. The updating formula for the variance of our belief $\bar{\sigma}_x^{2,n}$, assuming we measure $x^n = x$, is given by

$$\begin{aligned} \bar{\sigma}_x^{2,n} &= \left((\bar{\sigma}_x^{2,n-1})^{-1} + (\sigma_W^2)^{-1}\right)^{-1} \\ &= \frac{\bar{\sigma}_x^{2,n-1}}{1 + \bar{\sigma}_x^{2,n-1}/\sigma_W^2}. \qquad (4.7) \end{aligned}$$

Now let *Var*$^n(\cdot)$ be the variance of a random variable given what we know about the first n experiments. For example, *Var*$^n\bar{\mu}_x^n = 0$ since, given the first n experiments, $\bar{\mu}_x^n$ is deterministic. Next we compute the change in the variance in our belief about $\bar{\mu}^{n+1}$ given $\bar{\mu}^n$, given by

$$\tilde{\sigma}_x^{2,n} = Var^n[\bar{\mu}_x^{n+1} - \bar{\mu}_x^n].$$

We need to remember that given what we know at iteration n, which means given $\bar{\mu}^n$, the only reason that $\bar{\mu}^{n+1}$ is random (that is, with a variance that is not equal to zero) is because we have not yet observed W^{n+1}. As in Chapter 2, after some derivation, we can show that

$$\begin{aligned} \tilde{\sigma}_x^{2,n} &= \bar{\sigma}_x^{2,n} - \bar{\sigma}_x^{2,n+1} \qquad (4.8) \\ &= \frac{\bar{\sigma}_x^{2,n}}{1 + \sigma_W^2/\bar{\sigma}_x^{2,n}} \qquad (4.9) \\ &= (\beta_x^n)^{-1} - (\beta_x^n + \beta^W)^{-1}. \qquad (4.10) \end{aligned}$$

It is useful to compare the updating equation for the variance (4.7) with the change in the variance in (4.9). The formulas have a surprising symmetry to them. Equation (4.10) gives the expression in terms of the precisions.

We then compute ζ_x^n which is given by

$$\zeta_x^n = -\left|\frac{\bar{\mu}_x^n - \max_{x' \neq x} \bar{\mu}_{x'}^n}{\tilde{\sigma}_x^n}\right|. \qquad (4.11)$$

ζ_x^n is the *normalized influence* of decision x. It is the number of standard deviations from the current estimate of the value of decision x, given by $\bar{\mu}_x^n$, and the best alternative other than decision x. We always need to keep in mind that the value of information lies in its ability to change our decision (we first saw this in our decision tree illustration in section 1.6). So, we are always comparing the value of a choice to the best of all the other alternatives. The quantity ζ_x^n captures the distance between a choice and the next best alternative, measured in units of standard deviations of the change resulting from an experiment.

We next compute

$$f(\zeta) = \zeta\Phi(\zeta) + \phi(\zeta), \tag{4.12}$$

where $\Phi(\zeta)$ and $\phi(\zeta)$ are, respectively, the cumulative standard normal distribution and the standard normal density. That is,

$$\phi(\zeta) = \frac{1}{\sqrt{2\pi}}e^{-\frac{\zeta^2}{2}},$$

and

$$\Phi(\zeta) = \int_{-\infty}^{\zeta} \phi(x)dx.$$

The density $\phi(\zeta)$ is, of course, quite easy to compute. The cumulative distribution $\Phi(\zeta)$ cannot be calculated analytically, but very accurate approximations are easily available. For example, MATLAB provides the function `normcdf`(x, μ, σ), while Excel provides `NORMSDIST`(ζ). Searching the Internet for "calculate cumulative normal distribution" will also turn up analytical approximations of the cumulative normal distribution.

Finally, the knowledge gradient is given by

$$\nu_x^{KG,n} = \tilde{\sigma}_x^n f(\zeta_x^n). \tag{4.13}$$

Table 4.1 illustrates the calculations for a problem with five choices. The priors $\bar{\mu}^n$ are shown in the second column, followed by the prior precision. The precision of the experiment is $\beta^W = 1$.

Choice	$\bar{\mu}^n$	β^n	β^{n+1}	$\tilde{\sigma}$	$max'_x x'$	ζ	$f(\zeta)$	ν_x^{KG}
1	20.0	0.0625	1.0625	3.8806	28	-2.0616	0.0072	0.0279
2	22.0	0.1111	1.1111	2.8460	28	-2.1082	0.0063	0.0180
3	24.0	0.0400	1.0400	4.9029	28	-0.8158	0.1169	0.5731
4	26.0	0.1111	1.1111	2.8460	28	-0.7027	0.1422	0.4048
5	28.0	0.0625	1.0625	3.8806	26	-0.5154	0.1931	0.7493

Table 4.1 Calculations illustrating the knowledge gradient index

Interestingly, the knowledge gradient formula in (4.13) is symmetric. This means that, if we are looking for the alternative with the lowest value (rather than the highest), we still have

$$\mathbb{E}\left[\min_{x'} \bar{\mu}_{x'}^n - \min_{x'} \bar{\mu}_{x'}^{n+1} \,|\, S^n, x^n = x\right] = \tilde{\sigma}_x^n f\left(\zeta_x^n\right),$$

with the only difference being that $\max_{x' \neq x} \bar{\mu}^n_{x'}$ is replaced by $\min_{x' \neq x} \bar{\mu}^n_{x'}$ in the definition of ζ^n_x. This symmetry is a consequence of our choice of a Gaussian learning model with Gaussian observations. Intuitively, the normal density is symmetric about its mean, and thus, whether we are looking for the largest or the smallest normal value, the computation consists of integrating over the tail probability of some normal distribution. This does not mean that alternative x has the exact same KG factor in both cases (we are changing the definition of ζ^n_x), but it does mean that the KG formula retains the same basic form and intuitive interpretation. Later on, we show that this does not hold when the learning model is not Gaussian.

4.2.2 Some properties of the knowledge gradient

Recall that we provided a mathematical framework for an optimal learning policy in Section 3.2.2. It is important to keep in mind that the knowledge gradient policy is not optimal, in that it is not guaranteed to be the best possible policy for collecting information for any budget N. But for ranking and selection problems, the knowledge gradient policy has some nice properties. These include

- Property 1: The knowledge gradient is always positive, $\nu^{KG,n}_x \geq 0$ for all x. Thus, if the knowledge gradient of an alternative is zero, that means we won't measure it.
- Property 2: The knowledge gradient policy is optimal (by construction) if we are going to make exactly one experiment.
- Property 3: If there are only two choices, the knowledge gradient policy is optimal for any experiment budget N.
- Property 4: If N is our experimental budget, the knowledge gradient policy is guaranteed to find the best alternative as N is allowed to be big enough. That is, if x^N is the solution we obtain after N experiments, and
$$x^* = \arg\max \mu_x$$
is the true best alternative, then $x^N \to x^*$ as $N \to \infty$. This property is known as asymptotic optimality.
- Property 5: There are many heuristic policies that are asymptotically optimal (for example, pure exploration, mixed exploration-exploitation, epsilon-greedy exploration and Boltzmann exploration). But none of these heuristic policies is myopically optimal. The knowledge gradient policy is the only pure policy (an alternative term would be to say it is the only stationary policy) that is both myopically and asymptotically optimal.
- Property 6: The knowledge gradient has no tunable algorithmic parameters. Heuristics such as the Boltzmann policy (θ^B in equation (3.13)) and interval estimation (θ^{IE} in equation (3.17)) have tunable algorithmic parameters. The knowledge gradient has no such parameters, but as with all Bayesian methods, assumes we have a prior. Later, we demonstrate settings where we do have to introduce tunable parameters.

The knowledge gradient is not an optimal policy for collecting information, but these properties suggest that it is generally going to work well. There are situations where it can work poorly, as we demonstrate in section 4.3 below. In addition, it is more complicated to compute, and for more general belief models, it can be much more complicated to compute. This is a general property of all value-of-information policies.

4.3 THE VALUE OF INFORMATION AND THE S-CURVE EFFECT

The knowledge gradient computes the marginal value of information. It is easy to expect that the marginal value of information declines as we do more experiments, but this is not always the case. Below, we show how to identify when this property is violated, in which case the value of a single experiment can be quite low. When this happens, the knowledge gradient as described above does not add much value. We then describe some methods for overcoming this limitation by modeling what happens when you consider repeating the same experiment multiple times.

4.3.1 The S-curve

What if we perform n_x observations of alternative x, rather than just a single experiment? In this section, we derive the value of n_x experiments to study the marginal value of information. Note that this can be viewed as finding the value of a single experiment with precision $n_x\beta^W$, so below we view this as a single, more accurate experiment.

As before, let $\bar{\mu}^0_x$ and β^0_x be the mean and precision of our prior distribution of belief about μ_x. Now let $\bar{\mu}^1_x$ and β^1_x be the updated mean and precision after measuring alternative x a total of n_x times in a row. As before, we let $\beta^W = 1/\sigma^2_W$ be the precision of a single experiment. This means that our updated precision after n_x observations of x is

$$\beta^1_x = \beta^0_x + n_x\beta^W.$$

In Section 2.3.1, we showed that

$$\tilde{\sigma}^{2,n} \quad = \quad \mathit{Var}^n[\bar{\mu}^{n+1} - \bar{\mu}^n],$$

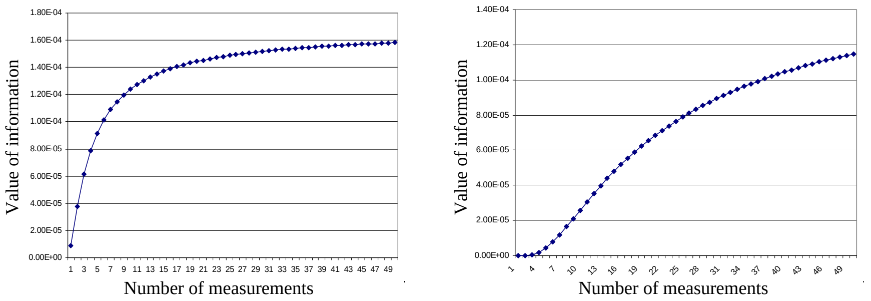
where Var^n is the conditional variance given what we know after n iterations. We are interested in the total variance reduction over n experiments. We denote this by $\tilde{\sigma}^{2,0}$, and calculate

$$\begin{aligned}\tilde{\sigma}^{2,0}(n_x) &= \bar{\sigma}^{2,0} - \bar{\sigma}^{2,1} \\ &= (\beta^0)^{-1} - (\beta^0 + n_x\beta^W)^{-1}.\end{aligned}$$

We next take advantage of the same steps we used to create equation (2.14) and write

$$\bar{\mu}^1_x = \bar{\mu}^0_x + \tilde{\sigma}^0_x(n_x)Z$$

where Z is a standard normal random variable, and where $\tilde{\sigma}^0_x(n_x) = \sqrt{\tilde{\sigma}^{2,0}_x(n_x)}$ is the standard deviation of the conditional change in the variance of $\bar{\mu}^1$ given that we make n_x observations.



4.2(a) 4.2(b)

Figure 4.2 Value of making n measures. In (a), the value of information is concave, while in (b) the value of information follows an S-curve.

We are now ready to calculate the value of our n_x experiments. Assume we are measuring a single alternative x, so $n_x > 0$ and $n_{x'} = 0$ for $x' \neq x$. Then we can write

$$\nu_x^{KG}(n_x) = \mathbb{E}\left[\max_{x'}(\bar{\mu}_{x'}^0 + \tilde{\sigma}_{x'}^0(n_{x'})Z_{x'})\right] - \max_{x'} \bar{\mu}_{x'}^0.$$

We can compute the value of n_x observations of alternative x using the knowledge gradient formula in equation (4.13),

$$\nu_x^{KG}(n_x) = \tilde{\sigma}_x^0(n_x) f\left(\frac{\bar{\mu}_x^0 - \max_{x' \neq x} \bar{\mu}_{x'}^0}{\tilde{\sigma}_x^0(n_x)}\right),$$

where $f(\zeta)$ is given in equation (4.12).

Now we have what we need to study some properties of the value of information. Consider a problem where $\sigma_W = 1.0$, $\sigma^0 = 1.5$ and $\Delta = \mu_x - \max_{x' \neq x} \mu_{x'} = 5$. Figure 4.2(a) shows the value of n experiments as n ranges from 1 to 50. This plot shows that the value of information is concave, as we might expect. Each additional experiment brings value, but less than the previous experiment, a behavior that seems quite intuitive.

Figure 4.2(b), however, gives the same plot for a problem where $\sigma_W = 2.5$. Note that when the experimental noise increases, the value of information forms an S-curve, with a very small value of information from the first few experiments, but then rising.

The S-curve behavior arises when a single experiment simply contains too little information to change a decision (and remember: the value of information is from its ability to change a decision). This behavior actually arises fairly frequently, and always arises when the outcome of an experiment is one of two outcomes such as success or failure (these are known as binomial outcomes). It is like introducing a new player to a team and then observing whether the team wins or not. You learn almost nothing about the new player purely from the outcome of whether or not the team won. In fact, it is entirely possible that the knowledge gradient may be some tiny number such as 10^{-10}.

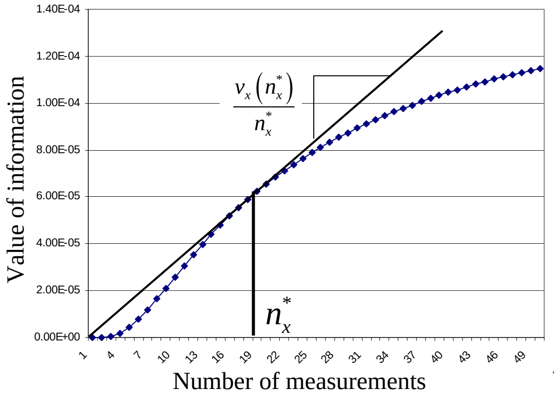


Figure 4.3 The KG(*) policy, which maximizes the average value of a series of experiments of a single alternative.

4.3.2 The KG(*) policy

A variant of the KG policy is called the KG(*) policy, which finds the number of experiments n_x^* which produces the highest *average* value of each observation, computed using

$$n_x^* = \arg\max_{n_x>0} \frac{\nu_x^{KG}(n_x)}{n_x}. \tag{4.14}$$

Figure 4.3 illustrates this policy. If the value of information is concave, then $n_x^* = 1$. If the function is non-concave, then it is very easy to find the value of n_x^* that solves equation (4.14).

4.3.3 A tunable lookahead policy

The KG(*) policy implicitly assumes that our experimental budget is large enough to sample alternative x roughly n_x^* times. There are many applications where this is just not going to be true. We can mimic the basic idea of KG(*), but replace n_x^* with a tunable parameter. Begin by defining

$\nu^{KG,n}(\theta^{KG}) =$ The knowledge gradient computed using a precision $\beta^1 = \beta^0 + \theta^{KG}\beta^W$.

We then define our tunable KG policy as

$$X^{KG}(\theta^{KG}) = \arg\max_x \nu_x^{KG,n}(\theta^{KG})$$

We now have to tune our policy to find the best value of θ^{KG} (see section 3.7 for our discussion on tuning). Here is where our policy adapts to our learning budget, as well as other problem characteristics that we choose to model.

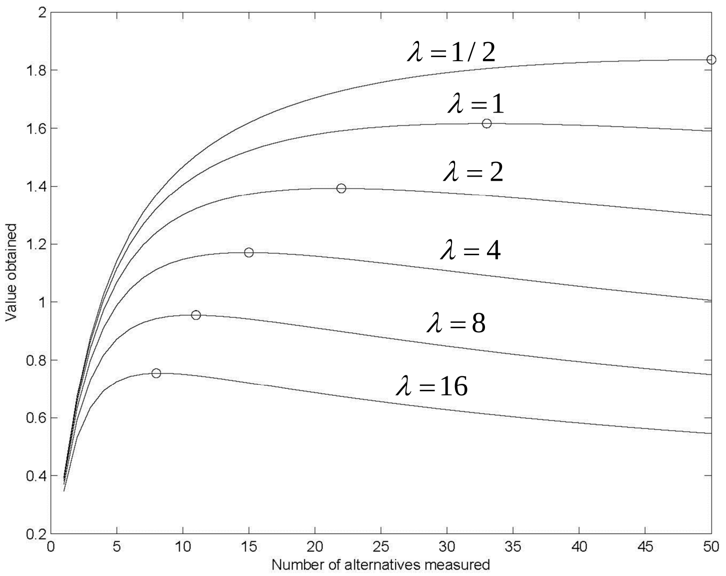


Figure 4.4 The value of spreading a budget of 50 experiments over $M \leq 50$ alternatives (from Frazier & Powell 2010).

4.3.4 The problem of too many choices

There are times when we simply have too many alternatives to evaluate. Imagine, for example, that we have 50 players to evaluate when forming a basketball team. We can let each take n_x shots and evaluate their shooting ability. Imagine that we have enough time to let all of them take a total of N shots. It is easy to imagine that if we have 50 players to consider, but only have enough time to let them take a total of 50 shots, we are not going to learn anything if each player gets to take a single shot.

We often find ourselves with too many choices. Consider the last time you visited a supermarket, or had to find a restaurant in a large city (even if you live there). Ultimately, we have to find a way to limit our choice set, and then find the best within that set.

Figure 4.4 illustrates the value of spreading a budget of 50 experiments uniformly over $N \leq 50$ alternatives. If the experimental noise λ is small (in the figure, this corresponds to $\lambda = 1/2$), we do the best if we can observe all 50 alternatives exactly once, after which we pick what appears to be the best choice. Since our evaluation is very accurate, we do the best when we have the most choices. As the experimental noise increases, we get more value if we focus our experimental budget over a subset of the alternatives. For example, if $\lambda = 16$, we do best if we arbitrarily choose eight alternatives out of our population of 50, and then focus on finding the best among these eight.

This behavior can help explain why some choices tend to be biased in favor of cosmetic differences. If you have to choose the best baseball player, there is a bias toward choosing tall, strong athletes, since these are attributes that are easy to identify and help to narrow the field. People are biased toward brands that are well known, and companies tend to work with established vendors. Simply landing near the top of an Internet search can also be a way of pruning a set of choices. However it is done, there are many examples where there is a need to use an initial filter to reduce the set of choices to one that can be evaluated using a fixed experimental budget.

4.4 THE FOUR DISTRIBUTIONS OF LEARNING

Perhaps one of the most subtle aspects of learning is understanding the different types of uncertainty, which depend on when you are asking the question. There are four different distributions that arise in the learning process. We illustrate these in the context of learning problems where all random variables are represented by normal distributions.

1) The *prior distribution* $N(\bar{\mu}^0_x, \bar{\sigma}^{2,0}_x)$, which gives our initial distribution of belief about the mean.

2) The *sampling distribution* of the random variable W, which is given by $N(\mu, \sigma^2_W)$, where μ is the true mean and σ^2_W is the sample variance.

3) The *posterior distribution* (after n experiments), given by $N(\bar{\mu}^n_x, \bar{\sigma}^{2,n}_x)$, which reflects the noise in the experimental outcome W. Here, we are modeling the distribution of μ after n experiments, but when we are at time $n = 0$ (that is, we have not run any experiments yet).

4) The conditional distribution of $\bar{\mu}^{n+1}_x$ given our beliefs at time n. This is also known as the *predictive distribution*. If we have seen $W^1, ..., W^n$, then $\bar{\mu}^{n+1}_x$ is random before we have observed W^{n+1}, which means that $\bar{\mu}^{n+1}_x \sim \mathcal{N}(\bar{\mu}^n_x, \tilde{\sigma}^{2,n}_x)$.

An understanding of these distributions is useful because it helps to highlight the different types of uncertainties we are trying to resolve by collecting information.

4.5 KNOWLEDGE GRADIENT FOR CORRELATED BELIEFS

There are many problems where updating our belief about one alternative tells us something about other alternatives. Some examples include

- We are trying to find the best set of features for a laptop. We try one laptop with 12G of RAM, a 2.8 Ghz processor, a 14” screen and weighing 3.2 pounds. We then offer a second laptop with 8G of RAM, but everything else the same. A third laptop has 10G of RAM, a 3.2 Ghz processor, a solid state internal disk drive, a 13” screen and weighs 2.5 lbs (with a much higher price). We start by selling the first laptop, and find that we are getting sales higher than expected. Given the similarities with the second laptop, it is reasonable to assume that the sales of this laptop will also be higher than expected.

- Now we are trying to find the best price for our laptop. We start with an initial guess of the sales volume we anticipate for prices in \$100 increments from \$800 to \$1500. We start at \$1100, and sales are much lower than we expected. This would lower our beliefs about sales at \$1000 and \$1200.

- We are trying to find the best of several paths to use for routing a bus. Each time the bus finishes a trip, we record the start and end times, but not the times for individual components of the trip (the bus driver does not have the time for this). If the travel time for one path is higher than expected, this would increase our estimates for other paths that have shared links.

- We are estimating who has a disease. If a concentration of a disease is higher than expected in one part of the population, then we would expect that people who are nearby, or otherwise have a reason to interact, will also have a higher likelihood of having the disease.

Correlations are particularly important when the number of possible experiments is much larger than the experimental budget. The experiment design might be continuous (where in the United States should we test birds for bird flu), or there may simply be a very large number of choices (such as websites relevant to a particular issue). The number of choices to measure may be far larger than our budget to measure them in a reliable way.

We begin our discussion by reviewing the updated formulas for correlated beliefs, first presented in section 2.3.2. We then describe how to compute the knowledge gradient using correlated beliefs, and close with a discussion of the benefits.

4.5.1 Updating with correlated beliefs

As we pointed out in section 3.5, we can handle correlations with any policy if we use this structure to update beliefs *after* we have decided which experiment to make. However, the knowledge gradient is able to imbed our knowledge of correlated beliefs when we compute the value of information.

We assume that we have a covariance matrix (or function) that tells us how experiments of x and x' are correlated. If x is a scalar, we might assume that the covariance of μ_x and $\mu_{x'}$ is given by

$$Cov(x, x') = \sigma_x \sigma_{x'} e^{-\beta|x-x'|},$$

where σ_x and $\sigma_{x'}$ is the standard deviation of our belief about μ_x and $\mu_{x'}$, respectively. Here, β controls the relationship between μ_x and $\mu_{x'}$ as the distance between x and x' changes. Or, we just assume that there is a known covariance matrix Σ with element $\sigma_{xx'}$. For now, we continue to assume that the alternatives x are discrete, an assumption we relax later.

There is a way of updating our estimate of $\bar{\mu}^n$ which gives us a more convenient analytical form than what is given in Section 2.3.2. To simplify the algebra a bit, we let $\lambda^W = \sigma_W^2 = 1/\beta^W$. As we did in Section 4.2, we need the *change* in the variance of our belief due to the results of an experiment. Following the development in Chapter 2, let $\Sigma^{n+1}(x)$ be the updated covariance matrix given that we have chosen to measure alternative x, and let $\tilde{\Sigma}^n(x)$ be the change in the covariance matrix due to evaluating x, which is given by

$$\begin{aligned} \tilde{\Sigma}^n(x) &= \Sigma^n - \Sigma^{n+1} && (4.15)\\ &= \frac{\Sigma^n e_x (e_x)^T \Sigma^n}{\Sigma^n_{xx} + \lambda^W}, && (4.16) \end{aligned}$$

where e_x is a column vector of 0's with a 1 in the position corresponding to x. We note that (4.15) closely parallels (4.8) for the case of independent beliefs. Now define the column vector $\tilde{\sigma}^n(x)$, which gives the change in our belief about each alternative x' resulting from measuring alternative x. This is calculated using

$$\tilde{\sigma}^n(x) = \frac{\Sigma^n e_x}{\sqrt{\Sigma^n_{xx} + \lambda^W}}. \qquad (4.17)$$

Keep in mind that $\tilde{\sigma}^n(x)$ is a vector with element $\tilde{\sigma}^n_i(x)$ for a given experiment $x = x^n$.

Also let $\tilde{\sigma}_i(x)$ be the component $(e_i)^T \tilde{\sigma}(x)$ of the vector $\tilde{\sigma}(x)$. Let $\mathit{Var}^n(\cdot)$ be the variance given what we know after n experiments. We note that if we measure alternative x^n, then

$$\begin{aligned} \mathit{Var}^n\left[W^{n+1}_{x^n} - \bar{\mu}^n_{x^n}\right] &= \mathit{Var}^n\left[\mu_{x^n} + \varepsilon^{n+1} - \bar{\mu}^n_{x^n}\right] && (4.18)\\ &= \mathit{Var}^n\left[\mu_{x^n} + \varepsilon^{n+1}\right] && (4.19)\\ &= \Sigma^n_{x^n x^n} + \lambda^W. && (4.20) \end{aligned}$$

We obtain equation (4.18) using $W^{n+1}_{x^n} = \mu_{x^n} + \varepsilon^{n+1}$, since $W^{n+1}_{x^n}$ is a noisy observation of the truth μ_{x^n} (which is random because of our Bayesian belief model). We drop $\bar{\mu}^n_{x^n}$ to obtain (4.18) because we are conditioning on what we know after n experiments, which means that $\bar{\mu}^n_{x^n}$ is deterministic (that is, it is just a number).

Next define the random variable Z^{n+1}

$$Z^{n+1} = \frac{(W^{n+1}_x - \bar{\mu}^n_x)}{\sqrt{\mathit{Var}^n\left[W^{n+1}_x - \bar{\mu}^n_x\right]}},$$

where $x = x^n$. The random variable W^{n+1}_x is normally distributed with mean $\bar{\mu}^n_x$ and variance $\mathit{Var}^n\left[W^{n+1}_x - \bar{\mu}^n_x\right] = \mathit{Var}^n\left[W^{n+1}_x\right]$, which means that Z^{n+1} is normally distributed with mean 0 and variance 1.

We use the notation $\bar{\mu}^{n+1}(x^n)$ to represent the updated vector of estimates of the mean $\bar{\mu}^n$, with element $\bar{\mu}^{n+1}_i(x^n)$ after running experiment x^n. The updating equation for correlated beliefs (first presented in section 2.3.2) is given by

$$\bar{\mu}^{n+1}(x^n) = \bar{\mu}^n + \frac{W^{n+1}_{x^n} - \bar{\mu}^n_{x^n}}{\lambda^W + \Sigma^n_{x^n x^n}} \Sigma^n e_{x^n}. \qquad (4.21)$$

We can rewrite (4.21) as

$$\bar{\mu}^{n+1}(x^n) = \bar{\mu}^n_{x^n} + \tilde{\sigma}^n(x^n) Z^{n+1}, \qquad (4.22)$$

where $\bar{\mu}^n_{x^n}$ and $\tilde{\sigma}^n(x^n)$ are vectors, while Z^{n+1} is a scalar.

Equation (4.22) nicely brings out the definition of $\tilde{\sigma}^n(x^n)$ as the vector of variances of the future estimates $\bar{\mu}^{n+1}_{x^n}$ given what we know at time n (which makes $\bar{\mu}^n$ deterministic). Essentially, even though a single experiment may change our beliefs about every alternative, the "randomness" in this change comes from a scalar observation, so the conditional distribution is expressed in terms of the scalar Z^{n+1}. This is a useful way of representing $\bar{\mu}^{n+1}_{x^n}$, especially for problems with correlated beliefs, as we see next.

4.5.2 Computing the knowledge gradient

We now face the challenge of computed the knowledge gradient. The knowledge gradient policy for correlated beliefs is computed using

$$\begin{aligned} X^{KG}(s) &= \arg\max_x \mathbb{E}\left[\max_i \bar{\mu}^{n+1}_i(x^n) \mid S^n = s, x^n = x\right] && (4.23) \\ &= \arg\max_x \mathbb{E}\left[\max_i \left(\bar{\mu}^n_i + \tilde{\sigma}^n_i(x^n) Z^{n+1}\right) \mid S^n, x^n = x\right]. \end{aligned}$$

where Z is a one-dimensional standard normal random variable. The problem with this expression is that the expectation is hard to compute. We encountered the same expectation when experiments are independent, but in this case we just have to do an easy computation involving the normal distribution. When the experiments are correlated, the calculation becomes more difficult.

There is a way to compute the expectation exactly. We start by defining

$$h(\bar{\mu}^n, \tilde{\sigma}^n(x)) = \mathbb{E}\left[\max_i \left(\bar{\mu}^n_i + \tilde{\sigma}^n_i(x^n) Z^{n+1}\right) \mid S^n, x^n = x\right]. \qquad (4.24)$$

Substituting (4.24) into (4.23) gives us

$$X^{KG}(s) = \arg\max_x h(\bar{\mu}^n, \tilde{\sigma}^n(x)). \qquad (4.25)$$

Now let $h(a, b) = \mathbb{E}\max_i(a_i + b_i Z)$, where $a = \bar{\mu}^n_i$, $b = \tilde{\sigma}^n_i(x^n)$ and Z is our standard normal deviate. Both a and b are M-dimensional vectors. We next sort the elements b_i so that $b_1 \leq b_2 \leq \ldots$ so that we get a sequence of lines with increasing slopes. If we plot the lines, we get a series of cuts similar to what is depicted in Figure 4.5. We see there are ranges for z over which the line for alternative 1 is higher than any other line; there is another range for z over which alternative 2 dominates. But there is no range over which alternative 3 dominates; in fact, alternative 3 is dominated by every other alternative for any value of z. What we need to do is to identify these dominated alternatives and drop them from further consideration.

To do this we start by finding the points where the lines intersect. If we consider the lines $a_i + b_i z$ and $a_{i+1} + b_{i+1} z$, we find they intersect at

$$z = c_i = \frac{a_i - a_{i+1}}{b_{i+1} - b_i}.$$

For the moment, we are going to assume that $b_{i+1} > b_i$ (that is, no ties). If $c_{i-1} < c_i < c_{i+1}$, then we can generally find a range for z over which a particular

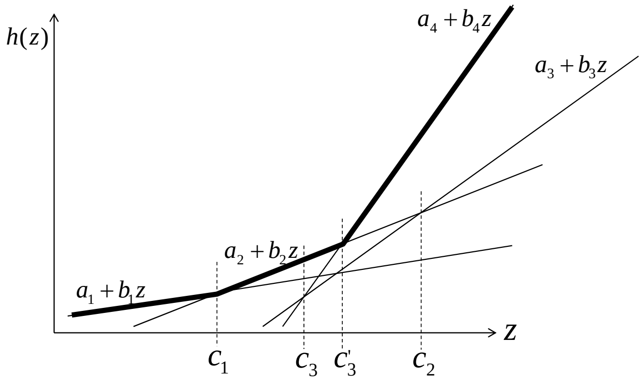


Figure 4.5 Regions of z over which different choices dominate. Choice 3 is always dominated.

choice dominates, as depicted in Figure 4.5. We can identify dominated alternatives such as alternative 3 in the figure when $c_{i+1} < c_i$. When this happens, we simply drop it from the set. Thus, instead of using c_2, we would use c'_3 to capture the intersection between the lines for choices 2 and 4.

Once we have the sequence c_i in hand, we can compute (4.23) using

$$h(a, b) = \sum_{i=1}^{M} (b_{i+1} - b_i) f(-|c_i|),$$

where as before, $f(z) = z\Phi(z) + \phi(z)$. Of course, the summation has to be adjusted to skip any choices i that were found to be dominated.

Figure 4.6 illustrates the use of the correlated knowledge gradient algorithm when it is applied to a 2-dimensional surface represented using a lookup table with correlated beliefs to capture the smoothness of the function. The figure shows the function after one experiment (figure 4.6(a)), four experiments (b), five experiments (c), and six experiments (d). These plots help to illustrates two behaviors: First, how the knowledge gradient tends to pursue regions that are farthest from previous experiments (where the uncertainty is higher), and second, the ability to generalize the results of experiment using correlated beliefs.

4.5.3 The value of a KG policy with correlated beliefs

It is clear that we should use the structure of the correlations to update our beliefs after an experiment, leaving the question: how much value do we obtain by incorporating correlations within the knowledge gradient calculation? The knowledge gradient with independent beliefs is much easier to calculate, so it would be nice if we could use the simpler policy but still use correlations when updating beliefs.

Figure 4.7 addresses this question by plotting the knowledge gradient for two policies: the knowledge gradient using independent beliefs, and the knowledge gradient

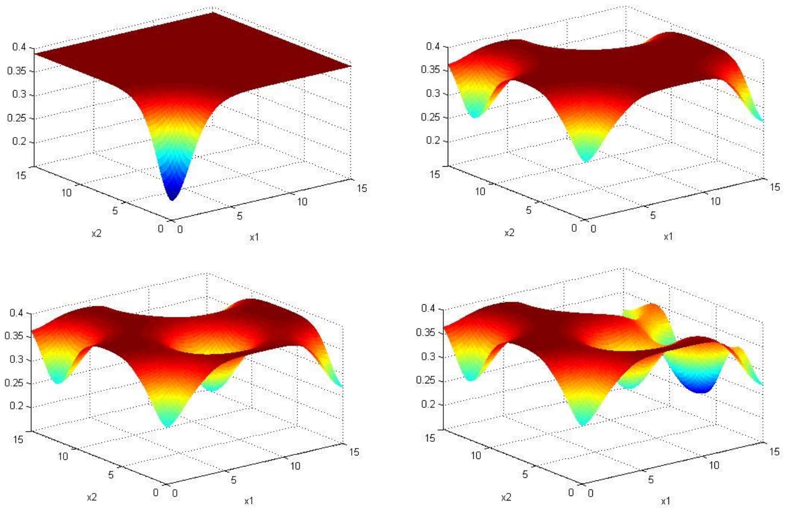


Figure 4.6 Evolution of belief about a 2-dimensional surface, showing the effect of experiments on our belief about the behavior of other experiments in the same proximity.

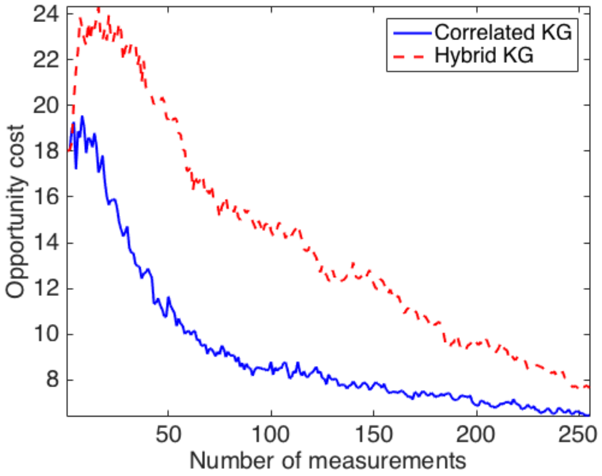


Figure 4.7 Performance of the knowledge gradient using independent or correlated beliefs when estimating the value of information.

incorporating correlated beliefs in the lookahead policy. Both policies are simulated using correlations when updating beliefs in the simulator, so this is a pure measure of the value of incorporating correlations when estimating the value of information. This plot suggests that there is considerable value capturing correlations within the policy, suggesting that the assumption of independence when estimating the value of information is not going to be efective.

Handling correlations between experiments has tremendous practical value. When we assumed independent experiments, it was necessary to measure each option at least once. There are applications where the number of potential experiments is far greater than the number of experiments that we can actually make. If we have information about correlations, we can find good designs even when the experimental budget is much smaller than the number of alternatives.

4.6 ANTICIPATORY VS. EXPERIENTIAL LEARNING

In section 3.5, we made the point that no matter what the policy does, we can capture complex belief models, along with complex dynamics of the information process, when we step forward from experiment n to $n+1$. Thus, all of the heuristic policies introduced in section 3.2 can capture correlated beliefs as part of the learning process.

Then, we just saw that we could include correlated beliefs in our model of the future as part of the process of deciding what experiment to run.

These are two very different ways of incorporating correlated beliefs (or, for that matter, any type of belief model). We distinguish these two modes using the following terms:

Anticipatory learning - This is the learning that we *anticipate* doing as part of our decision rule, when we are simulating decisions in the future.

Experiential learning - This is learning that we would do as we *experience* data from real observations from the field.

Anticipatory learning occurs within the decision function $X^{\pi}(S_t)$ when decisions require calculations that approximate in some way how new information may change our underlying belief. In our discussion of decision trees in Section 1.6, Section 1.6.1 presented a basic decision tree where outcomes did not change downstream probabilities. This is a case of a policy that does not use anticipatory learning. Sections 1.6.2 and 1.6.3 provided examples where observations of information were used to change downstream distributions within the decision tree used to make the decision. These sections provide examples of policies that use anticipatory learning.

Experiential learning is what happens after a decision has been made, and we then observe an actual outcome. Section 3.2.2 introduced the transition function $S_{t+1} = S^M(S_t, x_t, W_{t+1})$ which describes the updated belief state given an observation W_{t+1}. In some cases, W_{t+1} comes from a real physical system where the new information might be observed sales, or an observed response to a drug treatment. It is entirely possible that the distribution generating observed values of W_{t+1} is quite different than a distribution $f_W(w)$ that we are assuming for W_{t+1} within our policy $X^{\pi}(S_t)$. If this is the case, then it is natural to use observations of W_{t+1} to update our estimate of the density $f_W(w)$. This is experiential learning, and it is assumed to be part of the transition function $S^M(\cdot)$.

There are, however, many applications where $S^M(\cdot)$ represents a computer simulation, and W_{t+1} is generated from an assumed probability distribution $f_W(w)$. If

$f_W(w)$ is the distribution we use within our policy $X^\pi(S_t)$, then we are not actually learning anything from an observation W_{t+1} that comes from $f_W(w)$. This is quite common in many stochastic simulations, where the focus is on the uncertainty in W, and not in the uncertainty about the distribution $f_W(w)$ of W. An exception would be a simulation that is focusing explicitly on policies for collecting information. In such simulations, it is natural to make observations of W_{t+1} as we step forward using $S_{t+1} = S^M(S_t, x_t, W_{t+1})$ which are drawn from a truth that is unknown to the policy $X^\pi(S_t)$.

With these concepts in mind, we can describe three modeling strategies based on how new information is used.

1) No anticipatory or experiential learning. This is the default strategy in both deterministic and stochastic optimization, where we assume that observations of random variables are realizations from a known distribution, and as a result would not change our belief about the distribution. This situation is analogous to the “independent KG” curve in Figure 4.7.

2) Experiential learning without anticipatory learning. Here we use actual observations of random variables to update our beliefs, but we ignore the fact that we are going to do this when we make a decision. Learning is thus reduced to a statistical problem, with no optimization component. This is exactly the “hybrid KG” situation in Figure 4.7.

3) Experiential learning with anticipatory learning. This is where we use learning within the policy, as in equation (4.25), and in the transition function, as in equations (2.18)-(2.19). Such a situation corresponds to the “correlated KG” curve in Figure 4.7.

We ignore the combination of using anticipatory learning without experiential learning, because it does not make sense to anticipate the effect of observations on beliefs, but then not use real observations to update beliefs.

Anticipatory learning spans any policy (optimal or heuristic) which uses some representation of the uncertainty in our belief about the value of each choice. We did this heuristically in Chapter 3 with policies such as mixed exploitation-exploration, epsilon-greedy exploration, and Boltzmann exploration. In this chapter, the knowledge gradient, expected improvement and linear loss policies all represent methods which use some form of explicit learning within the learning policy.

4.7 THE KNOWLEDGE GRADIENT FOR SOME NON-GAUSSIAN DISTRIBUTIONS

It is important to remember that, while the definition of the knowledge gradient given in (4.3) is general, and can be written down for almost any learning problem, the KG formula given in (4.13) is geared specifically to the ranking and selection problem with independent alternatives and a normal-normal learning model (normally distributed prior, normally distributed experimental outcomes). This is not particularly surprising, because the formula uses the functions ϕ and Φ that characterize a

Gaussian distribution. We might expect other learning models to yield completely different formulas.

In this section, we give expressions for the marginal value of a single experiment for some of the non-Gaussian learning models presented in Chapter 2. They have certain notable differences from the normal-normal case, but all of them are based on the relationship between our beliefs about alternative x, and our beliefs about "the best alternative other than x." In all of these examples, we assume that the alternatives are independent, for the simple reason that there are no convenient Bayesian conjugate priors for any distribution other than normal.

4.7.1 The beta-Bernoulli model

We begin with the beta-Bernoulli model both for its simplicity which captures problems where outcomes are binary (e.g. success/failure, purchase a product or not) and where the uncertainty is the probability of success. This model will also highlight an important limitation of the basic knowledge gradient which tries to learn from a single outcome.

A popular application of the beta-bernoulli model arises in the setting of clinical trials. Our alternatives may correspond to different experimental medical treatments. A treatment can result in either success or failure, and we are interested in finding the treatment with the highest success probability. The challenge comes when we have a relatively small number of clinical trials that we can perform. We must decide which treatments to test in order to find the highest success probability most efficiently.

As in Chapter 2, we assume that the success probability ρ_x of treatment x follows a beta distribution with parameters α_x^0 and β_x^0. If we decide to test treatment x^n at time n, the result $W_{x^n}^{n+1}$ of the test is 1 with probability ρ_{x^n} and 0 with probability $1 - \rho_{x^n}$. Our beliefs about the treatment are updated using the equations

$$\alpha_x^{n+1} = \begin{cases} \alpha_x^n + W_x^{n+1} & \text{if } x^n = x \\ \alpha_x^n & \text{otherwise,} \end{cases}$$

$$\beta_x^{n+1} = \begin{cases} \beta_x^n + \left(1 - W_x^{n+1}\right) & \text{if } x^n = x \\ \beta_x^n & \text{otherwise.} \end{cases}$$

Our estimate of ρ_x given S^n is $\mathbb{E}\left(\rho_x \,|\, S^n\right) = \frac{\alpha_x^n}{\alpha_x^n + \beta_x^n}$. Recall that α_x^n roughly corresponds to the number of successes that we have observed in n trials, whereas β_x^n represents the number of failures. As usual, the KG factor is defined as

$$\nu_x^{KG,n} = \mathbb{E}\left[\max_{x'} \frac{\alpha_{x'}^{n+1}}{\alpha_{x'}^{n+1} + \beta_{x'}^{n+1}} - \max_{x'} \frac{\alpha_{x'}^n}{\alpha_{x'}^n + \beta_{x'}^n} \;\middle|\; S^n\right].$$

Unlike some of the other models discussed in this section, the predictive distribution in the beta-Bernoulli model is very simple. Given S^n and $x^n = x$, the conditional distribution of α_x^{n+1} is essentially Bernoulli. There are only two possible outcomes, whose probabilities are given by

$$\mathbb{P}\left(\alpha_x^{n+1} = \alpha_x^n + 1\right) = \frac{\alpha_x^n}{\alpha_x^n + \beta_x^n}, \qquad \mathbb{P}\left(\alpha_x^{n+1} = \alpha_x^n\right) = \frac{\beta_x^n}{\alpha_x^n + \beta_x^n}.$$

Choice	α^n	β^n	$\frac{\alpha^n}{\alpha^n+\beta^n+1}$	$\frac{\alpha^n}{\alpha^n+\beta^n}$	$\frac{\alpha^n+1}{\alpha^n+\beta^n+1}$	C^n	KG
1	1	13	0.0667	0.0714	0.1333	0.8874	0
2	2	11	0.1429	0.1538	0.2143	0.8874	0
3	1	20	0.0455	0.0476	0.0909	0.8874	0
4	67	333	0.1671	0.1675	0.1696	0.8874	0
5	268	34	0.8845	0.8874	0.8878	0.1675	0

Table 4.2 Calculations of the KG formula for the beta-Bernoulli model

The predictive distribution of β_x^{n+1} is also Bernoulli, but the probabilities are reversed. That is, $\mathbb{P}\left(\beta_x^{n+1} = \beta_x^n + 1\right)$ is now $\frac{\beta_x^n}{\alpha_x^n+\beta_x^n}$.

Letting $C_x^n = \max_{x' \neq x} \frac{\alpha_{x'}^n}{\alpha_{x'}^n+\beta_{x'}^n}$ as usual, we can derive the KG formula

$$\nu_x^{KG,n} = \begin{cases} \frac{\alpha_x^n}{\alpha_x^n+\beta_x^n}\left(\frac{\alpha_x^n+1}{\alpha_x^n+\beta_x^n+1} - C_x^n\right) & \text{if } \frac{\alpha_x^n}{\alpha_x^n+\beta_x^n} \leq C_x^n < \frac{\alpha_x^n+1}{\alpha_x^n+\beta_x^n+1} \\ \frac{\beta_x^n}{\alpha_x^n+\beta_x^n}\left(C_x^n - \frac{\alpha_x^n}{\alpha_x^n+\beta_x^n+1}\right) & \text{if } \frac{\alpha_x^n}{\alpha_x^n+\beta_x^n+1} \leq C_x^n < \frac{\alpha_x^n}{\alpha_x^n+\beta_x^n} \\ 0 & \text{otherwise.} \end{cases} \quad (4.26)$$

Observe that

$$\frac{\alpha_x^n}{\alpha_x^n+\beta_x^n+1} \leq \frac{\alpha_x^n}{\alpha_x^n+\beta_x^n} \leq \frac{\alpha_x^n+1}{\alpha_x^n+\beta_x^n+1},$$

and the KG factor depends on where C_x^n falls in relation to these quantities. In this case, we do not measure x if we believe ρ_x to be too low or too high. We will only benefit from measuring x if our beliefs about ρ_x are reasonably close to our beliefs about the other success probabilities. There is a certain symmetry to this formula (both low and high estimates result in knowledge gradients of zero), and in fact, it has the same symmetric quality of the KG formula for the normal-normal model. If our objective is to find the smallest success probability rather than the largest, (4.26) remains the same; the only change is that we replace the maximum in the definition of C_x^n by a minimum.

An unexpected consequence of this structure is that it is entirely possible for all the KG factors to be zero in the beta-Bernoulli problem. Table 4.2 illustrates one possible instance where this might happen. Intuitively, it can occur when we are already quite certain about the solution to the problem. In the problem shown in the table, it seems clear that alternative 5 is the best, with $268 + 34 = 302$ trials yielding a very high success probability. Among the other alternatives, there is some competition between 2 and 4, but neither is close to alternative 5. As a result, our beliefs about 2 and 4 are too low for one experiment to change their standing with respect to 5. Similarly, our beliefs about alternative 5 are too high for one experiment to change its standing with respect to any of the others. Thus, all the KG factors are zero, which essentially means that the KG method does not really care which alternative to measure (we can choose any one at random).

Of course, it is conceivable (though it may seem unlikely) that alternative 5 is not really the best. For instance, if we could measure alternative 2 several hundred times, we might find that it actually has a higher success probability, and we were simply unlucky enough to observe 11 failures in the beginning. Unfortunately, the KG method only looks ahead one time step into the future, and thus is unable to consider this possibility. This hints at possible limitations of the knowledge gradient approach in a situation where the observations are discrete, similar to the S-curve effect we saw in Section 4.3.

4.7.2 The gamma-exponential model

Suppose that we have M exponentially distributed random variables with means $\lambda_1, ..., \lambda_M$, and we wish to find the one with the largest rate. For instance, we might be looking at a number of servers, with the goal of finding the one with the highest (fastest) service rate. Alternately, we might have a number of products, each with an exponential daily demand. In this setting, we might want to discover which product has the lowest demand (corresponding to the largest exponential rate), so that we might take this product out of production. A final example is the problem of network routing, where we have a number of possible routes for sending packets, and the objective is to find the route with the lowest ping time. In all of these problems, the observations (service time, daily demand, network latency) are positive, which means that the normal-normal model is not a good fit.

Instead, we use the gamma-exponential model (which we first saw in Section 2.6.1). We start by assuming that the rate λ_x of each alternative is uncertain and follows a gamma distribution with parameters a_x^0 and b_x^0. When we choose to measure alternative x^n at time n, we make a random observation $W_{x^n}^{n+1} \sim Exp\left(\lambda_{x^n}\right)$, and update our beliefs according to the equations

$$\begin{aligned} a_x^{n+1} &= \begin{cases} a_x^n + 1 & \text{if } x^n = x \\ a_x^n & \text{otherwise,} \end{cases} \\ b_x^{n+1} &= \begin{cases} b_x^n + W_x^{n+1} & \text{if } x^n = x \\ b_x^n & \text{otherwise.} \end{cases} \end{aligned}$$

We still have an independent ranking and selection problem, as in (3.2) and (3.3), but the specific updating mechanism for the alternative we measure is taken from the gamma-exponential model in Chapter 2.

Recall that, given the beliefs a_x^n and b_x^n, our estimate of λ_x is a_x^n / b_x^n, the mean of the gamma distribution. The KG factor of alternative x at time n is now given by

$$\nu_x^{KG,n} = \mathbb{E}\left[\max_{x'} \frac{a_{x'}^{n+1}}{b_{x'}^{n+1}} - \max_{x'} \frac{a_{x'}^n}{b_{x'}^n} \mid S^n\right]. \tag{4.27}$$

We omit the steps in the computation of this expectation; interested readers can follow the procedure in Section 4.12.1 and compute the formula for themselves. However, we point out one interesting detail. As in the normal-normal case, the computation of $\nu_x^{KG,n}$ requires us to find the conditional distribution, given S^n and $x^n = x$, of the parameter b_x^{n+1}. This is known as the *predictive distribution* of b_x^n. It is easy to see that

$$\begin{aligned}\mathbb{P}\left(b_x^{n+1} > y \,|\, S^n, x^n = x\right) &= \mathbb{P}\left(W^{n+1} > y - b_x^n \,|\, S^n, x^n = x\right)\\ &= \mathbb{E}\left[\mathbb{P}\left(W^{n+1} > y - b_x^n \,|\, \lambda_x\right) \,|\, S^n, x^n = x\right]\\ &= \mathbb{E}\left[e^{-\lambda_x\left(y-b_x^n\right)} \,|\, S^n, x^n = x\right]\\ &= \left(\frac{b_x^n}{y}\right)^{a_x^n}.\end{aligned}$$

We get from the second line to the third line by using the cumulative distribution (cdf) of the exponential distribution, since W_x^{n+1} is exponential given λ_x. The last line uses the moment-generating function of the gamma distribution. From this calculation, it follows that the density of b_x^{n+1} is given by

$$f(y) = \frac{a_x^n \left(b_x^n\right)^{a_x^n}}{y^{a_x^n+1}},$$

which is precisely the Pareto density with parameters a_x^n and b_x^n.

Once we have this fact, computing (4.27) is a matter of taking an expectation of a certain function over the Pareto density. Define $C_x^n = \max_{x' \neq x} \frac{a_x^n}{b_x^n}$. As before, this quantity represents the "best of the rest" of our estimates of the rates. We then compute

$$\tilde{\nu}_x^n = \frac{\left(b_x^n\right)^{a_x^n} \left(C_x^n\right)^{a_x^n+1}}{\left(a_x^n+1\right)^{a_x^n+1}} \tag{4.28}$$

which is a sort of baseline knowledge gradient. However, depending on certain conditions, we may subtract an additional penalty from this quantity when we compute the final KG formula.

It can be shown that the KG formula is given by

$$\nu_x^{KG,n} = \begin{cases} \tilde{\nu}_x^n & \text{if } x = \arg\max_{x'} \frac{a_{x'}^n}{b_{x'}^n} \\ \tilde{\nu}_x^n - \left(C_x^n - \frac{a_x^n}{b_x^n}\right) & \text{if } \frac{a_x^n+1}{b_x^n} > \max_{x'} \frac{a_{x'}^n}{b_{x'}^n} \\ 0 & \text{otherwise.} \end{cases} \tag{4.29}$$

This particular formula is skewed towards exploitation. If $x \neq \arg\max_{x'} \frac{a_{x'}^n}{b_{x'}^n}$, we subtract an additional value $C_x^n - \frac{a_x^n}{b_x^n}$ from the KG factor of alternative x. However, if $x = \arg\max_{x'} \frac{a_{x'}^n}{b_{x'}^n}$, there is no additional penalty. Furthermore, if our estimate of λ_x is low enough that $\frac{a_x^n+1}{b_x^n} \leq \max_{x'} \frac{a_{x'}^n}{b_{x'}^n}$, the KG factor is automatically zero. Since $\nu_x^{KG,n} \geq 0$ for all x, just like in the normal-normal case, this means that we won't even consider an alternative if our beliefs about it are too low. So, there is a more pronounced slant in favor of alternatives with high estimates than we saw in the normal-normal setting.

Table 4.3 illustrates this issue for a problem with five alternatives. Based on our beliefs, it seems that alternative 5 is the best. This alternative also has the highest KG factor. Although our estimate of λ_3 is very close to our estimate of λ_5 (they differ

Choice	a^n	b^n	a^n/b^n	C^n	Too low?	$\tilde{\nu}_x^n$	Penalty	Final KG
1	1.0	7.2161	0.1386	0.3676	yes			0
2	4.0	12.1753	0.3285	0.3676	no	0.0472	0.0390	0.0081
3	3.0	8.1802	0.3667	0.3676	no	0.0390	0.0008	0.0382
4	5.0	19.3574	0.2583	0.3676	yes			0
5	2.0	5.4413	0.3676	0.3667	no	0.0540	0	0.0541

Table 4.3 Calculations of the KG formula for the gamma-exponential model

by less than 0.001), the KG factor for alternative 5 is nearly 1.5 times larger than the KG factor for alternative 3. Of the three remaining alternatives, only one is believed to be good enough to warrant a (very small) non-zero KG factor.

One should not take this numerical example too close to heart, however. If we keep the same numbers, but let $a_3^n = 1$ and $b_3^n = 2.78$, then alternative 3 will have the largest KG factor, even though our estimate of it would actually be 0.3597, lower than in Table 4.3. Just as in the normal-normal problem, the KG method weighs our estimate of a value against the variance of our beliefs. All other things being equal, alternatives with lower values of a_x^n (that is, alternatives that we have measured fewer times) tend to have higher KG factors, so the KG method will occasionally be moved to explore an alternative that does not currently seem to be the best.

As a final detail in our discussion, let us consider the case where our goal is to find the alternative with the smallest rate, rather than the largest. So, instead of looking for the product with the lowest demand (highest rate) in order to take it out of production, we are now looking for the product with the highest demand (lowest rate) in order to determine the most promising line of research for new products. In this case, the KG factor of alternative x is defined to be

$$\nu_x^{KG,n} = \mathbb{E}\left[\min_{x'} \frac{a_{x'}^n}{b_{x'}^n} - \min_{x'} \frac{a_{x'}^{n+1}}{b_{x'}^{n+1}} \mid S^n\right].$$

The predictive distribution of b_x^{n+1}, given S^n and $x^n = x$, is still Pareto with parameters a_x^n and b_x^n. As before, let $C_x^n = \min_{x' \neq x} \frac{a_{x'}^n}{b_{x'}^n}$, and let $\tilde{\nu}_x^n$ be as in (4.28). The KG formula resulting from taking the appropriate expectation over the Pareto density is

$$\nu_x^{KG,n} = \begin{cases} \tilde{\nu}_x^n & \text{if } x \neq \arg\max_{x'} \frac{a_{x'}^n}{b_{x'}^n} \\ \tilde{\nu}_x^n - \left(C_x^n - \frac{a_x^n}{b_x^n}\right) & \text{if } \frac{a_x^n+1}{b_x^n} > C_x^n \\ 0 & \text{otherwise.} \end{cases} \tag{4.30}$$

This formula is the mirror image of (4.29). It has a similar appearance, but its effect is precisely the opposite: it rewards exploration. In this case, we never impose a penalty on any alternative *except* for $\arg\max_{x'} \frac{a_{x'}^n}{b_{x'}^n}$. In fact, if our beliefs about the best alternative are too good (i.e. our estimate of the corresponding λ_x is too small), the KG factor of this alternative is zero, and we do not measure it. Even if we believe that the best alternative is relatively close to the second-best, we still penalize the one that we think is the best.

Recall that, in the normal-normal case, the KG formula was the same regardless if we were looking for the alternative with the largest or the smallest value. However, in the gamma-exponential model, there is a clear difference between minimizing and maximizing. The reason is because the gamma and exponential distributions are not symmetric. Our use of the gamma prior means that our estimate of λ_x could potentially take on any arbitrarily high value, but it can never go below zero. Thus, roughly speaking, our beliefs about λ_x are more likely to be low than to be high. To put it another way, the true value λ_x can always be higher than we think. Thus, if we are looking for the lowest value of λ_x, we need to push ourselves to explore more, so as not to get stuck on one alternative that seems to have a low rate. However, if we are looking for the highest value of λ_x, it is often enough to stick with an alternative that seems to have a high rate: if the rate is not really as high as we think, we will discover this quickly.

4.7.3 The gamma-Poisson model

Let us now consider a ranking and selection problem where our observations are discrete. For instance, we might consider the problem of finding the product with the highest average demand, assuming that the individual daily demands are integer-valued. Then, the daily demand for product x can be modeled as a Poisson random variable with rate λ_x. We assume that λ_x follows a gamma distribution with parameters a_x^0 and b_x^0. If we choose to measure the demand for product x^n after n days, our beliefs are updated according to the equations

$$\begin{aligned} a_x^{n+1} &= \begin{cases} a_x^n + N_x^{n+1} & \text{if } x^n = x \\ a_x^n & \text{otherwise,} \end{cases} \\ b_x^{n+1} &= \begin{cases} b_x^n + 1 & \text{if } x^n = x \\ b_x^n & \text{otherwise,} \end{cases} \end{aligned}$$

where N_x^{n+1} is the number of units of product x ordered on the next day. We assume that $N_x^{n+1} \sim Poisson\left(\lambda_x\right)$. If we are looking for the largest rate, the definition of the KG factor is once again

$$\nu_x^{KG,n} = \mathbb{E}\left[\max_{x'} \frac{a_{x'}^{n+1}}{b_{x'}^{n+1}} - \max_{x'} \frac{a_{x'}^n}{b_{x'}^n} \,\middle|\, S^n \right].$$

The problem looks deceptively similar to the gamma-exponential problem. However, the first difference is that the predictive distribution of a_x^{n+1} is now discrete, since our observation is Poisson. In fact, it can be shown that

$$\mathbb{P}\left(a_x^{n+1} = a_x^n + k \,|\, S^n, x^n = x\right) = \frac{\Gamma\left(a_x^n + k\right)}{\Gamma\left(a_x^n\right)\Gamma\left(k+1\right)} \left(\frac{b_x^n}{b_x^n + 1}\right)^{a_x^n} \left(\frac{1}{b_x^n + 1}\right)^k$$

for $k = 0, 1, 2, ...$ We can view this as a sort of generalization of the negative binomial distribution. In fact, if a_x^n is an integer, then

$$\frac{\Gamma\left(a_x^n + k\right)}{\Gamma\left(a_x^n\right)\Gamma\left(k+1\right)} = \binom{a_x^n + k - 1}{a_x^n - 1}$$

and the predictive distribution of a_x^{n+1} is the classic negative binomial distribution, with a_x^{n+1} representing the total number of Bernoulli trials that take place before a_x^n failures occur, with $\frac{1}{b_x^n+1}$ being the success probability.

There is no closed-form expression for the cdf of a negative binomial distribution; however, because the distribution is discrete, the cdf can always be evaluated exactly by computing and adding the appropriate terms of the probability mass function. Let $F_a(y) = \mathbb{P}(Y_a \leq y)$, where Y_a has the negative binomial distribution for a failures, with success probability $\frac{1}{b_x^n+1}$. The basic KG quantity equals

$$\tilde{\nu}_x^n = C_x^n F_{a_x^n}\left(C_x^n\left(b_x^n+1\right)\right) - \frac{a_x^n}{b_x^n} F_{a_x^n+1}\left(C_x^n\left(b_x^n+1\right)+1\right),$$

and it can be shown that the KG factor is given by

$$\nu_x^{KG,n} = \begin{cases} \tilde{\nu}_x^n - \left(C_x^n - \frac{a_x^n}{b_x^n}\right) & \text{if } x \neq \arg\max_{x'} \frac{a_{x'}^n}{b_{x'}^n} \\ \tilde{\nu}_x^n & \text{if } \frac{a_x^n}{b_x^n+1} \leq C_x^n \\ 0 & \text{otherwise.} \end{cases} \tag{4.31}$$

Interestingly, if our estimate of λ_x is too high, we will not measure x at all, but if we measure an alternative that does not seem to have the highest rate, the KG factor has an extra penalty.

4.7.4 The Pareto-uniform model

Let us consider the problem of finding the product with the largest demand from a different angle. Suppose now that the demand for product x is uniformly distributed on the interval $[0, B_x]$, where B_x is unknown. We assume that B_x follows a Pareto distribution with parameters $\alpha^0 > 1$ and $b^0 > 0$. When we choose to perform a market study on product x at time n, we update our beliefs about x according to

$$b_x^{n+1} = \begin{cases} \max\left(b_x^n, W_x^{n+1}\right) & \text{if } x^n = x \\ b_x^n & \text{otherwise,} \end{cases}$$

$$\alpha_x^{n+1} = \begin{cases} \alpha_x^n + 1 & \text{if } x^n = x \\ \alpha_x^n & \text{otherwise.} \end{cases}$$

The random variable W_x^{n+1} is our observation of the demand for product x, and is uniform on $[0, B_x]$.

The goal is to find the product with the largest possible demand $\max_x B_x$. Recall that $\mathbb{E}\left[B_x \mid S^n\right] = \frac{\alpha_x^n b_x^n}{\alpha_x^n - 1}$. Then, the knowledge gradient is defined to be

$$\nu_x^{KG,n} = \mathbb{E}\left[\max_{x'} \frac{\alpha_{x'}^{n+1} b_{x'}^{n+1}}{\alpha_{x'}^{n+1} - 1} - \max_{x'} \frac{\alpha_{x'}^n b_{x'}^n}{\alpha_{x'}^n - 1} \mid S^n\right].$$

The procedure for computing this expectation is the same as in the other models that we have considered. One thing that makes it a bit messier than usual is that the predictive distribution of b_x^{n+1}, given S^n and $x^n = x$, is neither discrete nor continuous, but a mixture of both. This is because, if we measure x, then b_x^{n+1} is the maximum of a constant and our observation, which has the effect of "folding" part of the continuous density of the observation. It can be shown that

$$\mathbb{P}\left(b_x^{n+1} = b_x^n \,|\, S^n, x^n = x\right) = \frac{\alpha_x^n}{\alpha_x^n + 1}.$$

For $y > b_x^n$, however, the predictive distribution has a scaled Pareto density

$$f\left(y \,|\, S^n, x^n = x\right) = \frac{1}{\alpha_x^n + 1} \frac{\alpha_x^n \left(b_x^n\right)^{\alpha_x^n}}{y^{\alpha_x^n + 1}}.$$

This mixed distribution complicates the computation slightly, but the principle is the same. As usual, let

$$C_x^n = \max_{x' \neq x} \frac{\alpha_{x'}^n b_{x'}^n}{\alpha_{x'}^n - 1}$$

denote the "best of the rest." In this case, the basic KG quantity is

$$\tilde{\nu}_x^n = \frac{1}{\alpha_x^n \left(\alpha_x^n - 1\right)} \frac{\left(\alpha_x^n + 1\right)^{\alpha_x^n - 1} \left(b_x^n\right)^{\alpha_x^n}}{\left(\alpha_x^n\right)^{\alpha_x^n - 1} \left(C_x^n\right)^{\alpha_x^n - 1}}$$

and the KG formula is

$$\nu_x^{KG,n} = \begin{cases} \tilde{\nu}_x^n & \text{if } x \neq \arg\max_{x'} \frac{\alpha_{x'}^n b_{x'}^n}{\alpha_{x'}^n - 1} \\ \tilde{\nu}_x^n - \left(\frac{\alpha_x^n b_x^n}{\alpha_x^n - 1} - C_x^n\right) & \text{if } \frac{(\alpha_x^n + 1) b_x^n}{\alpha_x^n} \leq C_x^n \\ 0 & \text{otherwise.} \end{cases} \tag{4.32}$$

Thus, if we think that x has the largest possible demand, and our estimate of B_x is too much larger than our other estimates, we do not measure x. If x seems to have the largest demand, but the actual estimate is relatively close to our other estimates, the KG factor of x is non-zero, but has an extra penalty term. Otherwise, the KG factor is equal to $\tilde{\nu}_x^n$, with no additional penalty. This particular model promotes exploration.

4.7.5 The normal distribution as an approximation

While it is interesting to see the knowledge gradient computed for different distributions, most of the time we are going to use the normal distribution as an approximation for both the prior and posterior, regardless of the distribution of W. The reason, which we first addressed in section 2.6.6, is the central limit theorem. When we combine estimates from even a relatively small number of experiments (possibly as low as five), the normal distribution becomes a very good approximation, even when our experimental outcomes W are highly non-normal (e.g. binary).

The research community recognizes the value of using an approximation for the posterior, and has given it the name of *assumed density filtering* (or ADF). This is nothing more than taking a posterior distribution (density), and replacing it with a simpler one (such as the normal), which is a kind of filter.

Sampling distribution	Prior distribution	Predictive distribution
Normal	Normal	Normal
Exponential	Gamma	Pareto
Poisson	Gamma	Negative binomial
Uniform	Pareto	Mixed discrete/Pareto
Bernoulli	Beta	Bernoulli
Multinomial	Dirichlet	Multinomial
Normal	Normal-gamma	Student's t-distribution

Table 4.4 Table showing the distinctions between sampling, prior and predictive distributions for different learning models.

4.7.6 Discussion

Our examination of ranking and selection with non-Gaussian learning models underscores several interesting issues that were not as clear in the basic normal-normal problem. For one thing, there is now a real distinction between the prior, sampling and predictive distributions. In the normal-normal model, all three of these distributions were normal, but in other problems, all three can come from different families.

So far, we have managed to derive knowledge gradient formulas in all of these cases. It is clear, however, that a knowledge gradient formula depends not only on the particular type of learning model that we are using, but also on the objective function. In the gamma-exponential model, the KG algorithm uses two different formulas depending on whether we are looking for the largest exponential parameter or the smallest. Even in the basic normal-normal model, we can change the KG formula completely by changing the objective function. Exercises 4.17 and 4.19 give two examples of problems where this occurs.

The power and appeal of the KG method come from our ability to write the definition of the knowledge gradient in (4.3) in terms of some arbitrary objective function. Every new objective function requires us to recompute the KG formula, but as long as we are able to do this, we can create algorithms for problems where the objective function is very complicated. This allows us to go beyond the simple framework of ranking and selection, where the goal is always to pick the alternative with the highest value. Later on in this book, we will create knowledge gradient methods for problems where the alternatives are viewed as components that make up a large system, and the objective is a complicated function of our beliefs about the alternatives that somehow expresses the value of the entire system.

4.8 EXPECTED IMPROVEMENT

The expected improvement (EI) policy grew out of a class of learning problems where, instead of dealing with a finite set of alternatives, we have a continuous spectrum. For example, a semiconductor manufacturer uses liquid argon to provide an environment for sputtering, or depositing thin layers of metal, on semiconductor wafers. Argon is expensive and has to be purchased from a chemical manufacturing company, so the exact amount x that is needed may require some fine-tuning. We study this problem class in much more detail in Chapter 14. For now, we can examine EI in the context of the basic ranking and selection problem with M alternatives.

EI defines improvement in the following way. At time n, our current estimate of the largest value is given by $\max_{x'} \bar{\mu}^n_{x'}$. We would like to find alternatives x whose true value μ_x is greater than this estimated quantity. To put it another way, we prefer x such that μ_x is more likely to exceed (or improve upon) our current estimate. We then define the EI factor of x as

$$\nu^{EI,n}_x = \mathbb{E}\left[\max\left\{\mu_x - \max_{x'} \bar{\mu}^n_{x'}, 0\right\} \,\middle|\, S^n, x^n = x\right]. \tag{4.33}$$

We can interpret the EI factor as the value of collecting a single observation about x vs. doing nothing. According to EI, collecting information about x is only valuable if $\mu_x > \max_{x'} \bar{\mu}^n_{x'}$, that is, if μ_x brings about an improvement. Otherwise, the value is zero. Since μ_x is unknown at time n (but $\max_{x'} \bar{\mu}^n_{x'}$ is known), we take an expected value of the improvement over the distribution of μ_x. At time n, this is $\mathcal{N}\left(\bar{\mu}^n_x, (\sigma^n_x)^2\right)$. Like KG, the EI expectation in (4.33) leads to an explicit formula

$$\nu^{EI,n}_x = \sigma^n_x f\left(\frac{\bar{\mu}^n_x - \max_{x'} \bar{\mu}^n_{x'}}{\sigma^n_x}\right), \tag{4.34}$$

which closely resembles (4.13), with some differences that we highlight below. The policy then makes the decision using

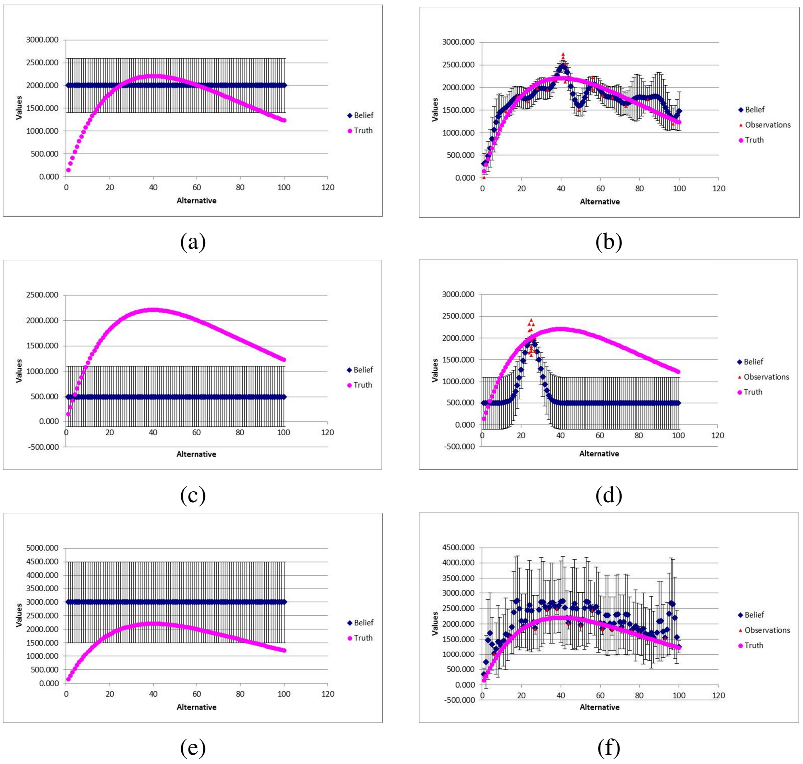
$$X^{EI,n} = \arg\max_x \nu^{EI,n}_x.$$

There are two main differences between EI and KG. First, $\nu^{KG,n}_x$ is calculated based on how likely $\bar{\mu}^{n+1}_x$ is to exceed $\max_{x' \neq x} \bar{\mu}^n_{x'}$; that is, x is compared to the best of the other alternatives. On the other hand, EI simply uses the current best estimate $\max_{x'} \bar{\mu}^n_{x'}$, the maximum over all alternatives, as the reference point. Second, EI uses the prior variance σ^n_x instead of the one-period variance reduction $\tilde{\sigma}^n_x$. One way to interpret this is that EI essentially assumes $\sigma^2_W = 0$. That is, EI considers what would happen if we could learn μ_x exactly in a single experiment.

We can still use EI in problems with noisy experiments, but the noise is not explicitly considered by the EI calculation. The literature on EI sometimes goes so far as to assume that $\sigma^2_W = 0$ in the underlying learning problem. This is not very interesting in ranking and selection, because we could find the exact optimal solution just by measuring every alternative once. However, in problems with continuous decisions, finding the best x is still challenging even when observations are exact. We return to EI in Chapter 14, where it is known as efficient global optimization or EGO.

4.9 THE PROBLEM OF PRIORS

Perhaps one of the most important yet difficult challenges in learning is constructing a reasonable prior. Sometimes a domain expert may have a reasonable understanding



(a) (b)

(c) (d)

(e) (f)

Figure 4.8 The effect of the prior on the search process. (a) shows an unbiased prior; (b) illustrates the resulting outcomes from the experiments. (c) shows a prior that is biased low; (d) illustrates that this produces a set of experiments focused around whatever point is chosen first. (e) shows a prior that is biased high; (f) illustrates that a high prior produces a search that evaluates the entire function.

about the shape of a function. The domain expert may refer to a person or to information derived from the Internet or prior studies. In the worst case, we may have to resort to doing an initial sample using randomly chosen observations. If we insist on using a Bayesian model, this last strategy is referred to as *empirical Bayes*.

The biggest problem arises when we have a prior, but it is not a very good one. Furthermore, we may be willing to assume that the precision of our prior belief is higher than it really is. Figure 4.8 presents a thought experiment with several possible scenarios. In Figure 4.8(a), we have an unknown function, a constant prior over the interval and a confidence interval that reasonably captures the spread of the function. This is a fairly accurate statement that we have no idea where the optimum is, but we do have an idea of the range of the function. Figure 4.8(b) is our estimate of the function after a series of experiments using the knowledge gradient, which shows that we do a very good job of finding the true optimum, with a very high level of confidence (note that we do not produce a precise estimate of the value of the function at the optimum).

Figure 4.8(c) illustrates a prior that is too low, and where we do not accurately represent the precision of our belief. 4.8(d) then shows that we start by sampling the function at a random point, and not surprisingly the observed value is much higher than our prior. As a result, we tend to focus our search near this point, since our prior suggests that there is no value in sampling points far from the initial sample. If we were to repeat the exercise with a different random starting point, we would have focused our entire search close to that point. The result is an unreliable estimate of the optimum.

Figure 4.8(e) then depicts a prior that is too high. When this is the case, Figure 4.8(f) shows that we end up sampling the entire function, since every observation produces an estimate that is very low relative to our prior estimate for the rest of the function. We then proceed to sample any point that has not been sampled before (and which is not close to a previously sampled point). This works if our budget is large enough to sample the entire function, but it would work poorly if this were not the case.

Perhaps the most visible lesson from this illustration is that it is important to be honest about the uncertainty in the prior. Figures 4.8(c) and 4.8(e) both display confidence intervals that do not cover the function. A narrow confidence bound is particularly problematic if the prior is biased low, because we tend to focus all of our energy around any point that we sample.

Of course, from the Bayesian point of view, there is no one fixed function. Rather, our prior encodes our beliefs about a wide range of possible true functions. The message of the preceding example can be taken to mean that we should make sure that this range is in some way “wide enough.”

Care should be used to ensure that the prior does not exclude portions of the function that may be quite good. For this reason, it is better to be biased high, and to use care not to overstate the precision of your prior (that is, where your confidence interval is too narrow). However, using a prior that is too high, and/or a confidence interval that is too large, may simply result in a lot of unnecessary exploration. In other words, the better your prior is, the better your search will be. This may be a way of saying that there is no free lunch in information collection.

4.10 VALUE OF INFORMATION WITH UNKNOWN VARIANCE*

An important simplifying assumption that we made for the knowledge gradient is that we know the variance (or precision) of an experimental outcome. In this section we consider the more difficult case of unknown variance, which was developed under the name “linear loss.” Not surprisingly, this is a more difficult problem, with a corresponding increase in complexity.

Two policies have been proposed in the literature under the names $LL(1)$ and LL_1, which might leave the impression that they are the same policy. Despite having very similar names, $LL(1)$ and LL_1 are actually two different policies. Both are examples of “linear-loss” or LL methods, close analogs to KG that were developed for ranking and selection with unknown means and variances. We discussed learning in this setting in Section 2.6.5.

Supposing that alternative x has value μ_x, and our observations of μ_x also have an unknown precision β^W_x, we can represent our beliefs about the mean and precision using a normal-gamma prior. Recall that, when we say that $\left(\mu_x, \beta^W_x\right)$ follows a normal-gamma distribution with parameters $\bar{\mu}^0_x, \tau^0_x, a^0_x, b^0_x$, we mean that β^W_x is gamma with parameters a^0_x and b^0_x, and the conditional distribution of μ_x given that $\beta^W = r$ is normal with mean $\bar{\mu}^0_x$ and precision $\tau^0_x r$. For each observation of μ_x that we collect, we use (2.42)-(2.45) to update our beliefs.

Both $LL(1)$ and LL_1 originate from a technique called $LL(N)$. This procedure was designed for a version of ranking and selection where, instead of collecting observations one at a time, we collect N of them at once, in a batch. Instead of making sequential decisions, where we first choose an alternative, then collect an observation, update our beliefs and make a new decision based on the new information, we make only one decision. Before collecting the batch of information, we choose how many observations (out of N total) should be collected for each alternative. For example, if we have three alternatives and $N = 10$, we might make a decision to collect $k_1 = 3$ observations of μ_1, $k_2 = 5$ observations of μ_2, and $k_3 = 2$ observations of μ_3. We make this decision based on the beliefs we have prior to collecting the batch. For example, if we have a lot of uncertainty about μ_2 (corresponding to a low value of τ^0_2), we might want to assign more observations to alternative 2.

Our goal is to choose an allocation vector $k = (k_1, ..., k_M)$ with $\sum_{i=1}^M k_i = N$ and $k_i \geq 0$ to maximize the usual objective function (**??**). In this setting, we can write the objective as

$$\max_k \mathbb{E} F^k \left(\mu, \beta^W, W\right) \tag{4.35}$$

with $F^k\left(\mu, \beta^W, W\right) = \max_x \bar{\mu}^n_x$. Notice that β^W is now included as part of the "truth," since it is also a vector of unknown parameters and affects the observations we get. The name "linear loss" is due to the fact that maximizing $\mathbb{E}F^k$ is the same as minimizing an equivalent expression $\mathbb{E}\left[\mathbb{E}^N\left(\max_x \mu_x - \max_x \bar{\mu}^n_x\right)\right]$, the expected loss or difference between the true best value and the estimated best value.

We cannot compute (4.35) directly, but we can approximate it. Let $x^* = \arg\max_x \bar{\mu}^0_x$ be the alternative that seems to be the best initially, before any information is collected. Because x^* does not depend on the choice of allocation k, we can rewrite (4.35) as

$$\max_k \mathbb{E}\left(\max_x \bar{\mu}^n_x - \bar{\mu}^n_{x^*}\right). \tag{4.36}$$

Although (4.35) and (4.36) have different optimal values, they have the same optimal solution (that is, the same k maximizes both functions). Define an indicator function

$$I^N_x = \begin{cases} 1 & \text{if } x = \arg\max_{x'} \bar{\mu}^n_{x'} \\ 0 & \text{otherwise,} \end{cases}$$

that equals 1 if and only if x is believed to be the best alternative after N experiments have been made. Also define a second indicator

$$\tilde{I}^N_x = \begin{cases} 1 & \text{if } \bar{\mu}^n_x \geq \bar{\mu}^n_{x^*} \\ 0 & \text{otherwise,} \end{cases}$$

to show whether x is believed to be better than x^* (but not necessarily better than all other alternatives) at time N. Then,

$$
\begin{aligned}
\mathbb{E}\left(\max_x \bar{\mu}_x^n - \bar{\mu}_{x^*}^n\right) &= \mathbb{E}\sum_{x=1}^{M} I_x^N \cdot \left(\bar{\mu}_x^n - \bar{\mu}_{x^*}^n\right) \\
&\leq \mathbb{E}\sum_{x=1}^{M} \tilde{I}_x^N \cdot \left(\bar{\mu}_x^n - \bar{\mu}_{x^*}^n\right) \\
&= \sum_{x=1}^{M} P\left(\bar{\mu}_x^n \geq \bar{\mu}_{x^*}^n\right) \mathbb{E}\left(\bar{\mu}_x^n - \bar{\mu}_{x^*}^n \mid \bar{\mu}_x^n \geq \bar{\mu}_{x^*}^n\right). \qquad (4.37)
\end{aligned}
$$

If we replace our objective function by the upper bound in (4.37), we have replaced the objective function $\max_x \bar{\mu}_x^n$ with a sum of pairwise comparisons between individual $\bar{\mu}_x^n$ and a single value $\bar{\mu}_{x^*}^n$. This is more tractable because we can express the predictive distribution of $\bar{\mu}_x^n$ at time 0, given that we will allocate k_x observations to alternative x, in terms of Student's t-distribution. Specifically, $\sqrt{\frac{\tau_x^0 a_x^0 (\tau_x^0 + k_x)}{b_x^0 k_x}} \left(\bar{\mu}_x^n - \bar{\mu}_x^0\right)$ follows a t-distribution with $2a_x^0$ degrees of freedom. We can then compute the expectation in (4.37) for a particular choice of k, and maximize k by setting the derivative with respect to k_x equal to zero. After a few more approximations, the LL method arrives at the allocation

$$
k_x = \frac{N + \sum_{x'=1}^{M} \tau_{x'}^0}{\sum_{x'=1}^{M} \sqrt{\frac{b_{x'}^0 \eta_{x'}^0}{b_x^0 \eta_x^0}}} - \tau_x^0 \qquad (4.38)
$$

where

$$
\eta_x^0 = \begin{cases} \sqrt{\lambda_{x,x^*}^0 \frac{2a_x^0 + \lambda_{x,x^*}^0 \left(\bar{\mu}_{x^*}^0 - \bar{\mu}_x^0\right)^2}{2a_x^0 - 1}} \phi_{2a_x^0}\left(\sqrt{\lambda_{x,x^*}^0}\left(\bar{\mu}_{x^*}^0 - \bar{\mu}_x^0\right)\right) & x \neq x^* \\ \sum_{x' \neq x^*} \eta_{x'}^0 & x = x^*, \end{cases}
$$

the function $\phi_{2\alpha}$ is the density of the t-distribution with 2α degrees of freedom, and

$$
\lambda_{x,x^*}^0 = \left(\frac{b_x^0}{\tau_x^0 a_x^0} + \frac{b_{x^*}^0}{\tau_{x^*}^0 a_{x^*}^0}\right)^{-1}. \qquad (4.39)
$$

The procedure uses several approximations, such as allowing k_x to be continuous when computing the optimal allocation, so it is possible for (4.38) to produce negative numbers. In this case, we need to adjust the allocation manually by rounding k_x up or down while ensuring that the allocations still add up to N.

What happens when we leave the batch setting and return to our usual world of sequential experiments? One easy solution is to run the $LL(N)$ method for $N = 1$ at every time step $n = 0, 1, 2, ..., N-1$. This is precisely what is meant by the $LL(1)$ technique. As a result, we will obtain a vector k^n of the form e_{x^n}, a vector of zeroes with a single 1 corresponding to the alternative that we need to measure. We then measure the alternative, update our beliefs, and repeat the same procedure assuming a batch of size 1.

The similarly-named LL_1 procedure works along the same lines. The main difference is that, in the batch setting, LL_1 assumes that all N samples will be allocated to a single alternative, and then finds the most suitable alternative under these conditions. This significantly simplifies computation, because the batch is guaranteed to change only a single set of beliefs. Setting $N = 1$ inside the policy then produces a straightforward rule for making decisions at time n,

$$X^{LL_1,n} = \arg\max_x \nu_x^{LL_1,n},$$

where

$$\nu_x^{LL_1,n} = \sqrt{\lambda_{x,\tilde{x}}^n}\Psi_{2a_x^n}\left(\sqrt{\lambda_{x,\tilde{x}}^n}\left|\bar{\mu}_x^n - \bar{\mu}_{\tilde{x}}^n\right|\right), \tag{4.40}$$

for $\tilde{x} = \arg\max_{x' \neq x} \bar{\mu}_x^n$, and $\Psi_d(z) = \frac{d+z^2}{d-1}\phi_d(z) - z\Phi_d(-z)$, with ϕ_d, Φ_d being the pdf and cdf of the standard t-distribution with d degrees of freedom. The quantity $\lambda_{x,\tilde{x}}^n$ is computed exactly as in (4.39), but using the time-n beliefs. Notice that this procedure compares our beliefs about x to our beliefs about $\tilde{x}$, the best alternative other than x. This recalls the behavior of the KG policy, and indeed, LL_1 is the exact analog of KG for ranking and selection with unknown precision in the noise of an experiment.

The computation required for LL_1 is simpler and requires fewer approximations than $LL(1)$. It is less suitable for the batch setting, where we would most likely choose to divide our observations among many alternatives, but better suited to the sequential setting considered in this chapter. Additionally, some experiments in the simulation literature suggest that we can learn more effectively and discover better alternatives by using LL_1 to make decisions sequentially or with very small batches, rather than by running $LL(N)$ to allocate our entire learning budget at once.

4.11 DISCUSSION

The appeal of the knowledge gradient is that it is a simple idea that can be applied to many settings. A particularly powerful feature of the knowledge gradient is that it can capture the important dimension of correlated beliefs. In fact, it is useful to review the list of applications given in Section 1.2 where you will see that almost all of these are characterized by correlated beliefs. Later in the volume, we consider more complex sets of alternatives such as finding the best subset, or the best value of a continuous, multidimensional parameter vector.

The knowledge gradient (as with any policy based on the expected value of a single experiment) is vulnerable to the nonconcavity of information. Indeed, if you have a problem where the value of information is nonconcave, then you have to address the issues discussed in Section 4.3, regardless of your choice of learning policy. However, if the value of a single experiment is nonconcave, then this simply means that you have to think about taking repeated experiments. This behavior would almost always be present if the information W^n is binomial, which means that we should be thinking about the value of multiple trials.

We have presented results for situations other than the normal-normal model, but we suspect that most applications will lend themselves reasonably well to the normal-normal model, for two reasons. First, while the initial prior may not be normal, the central limit theorem generally means that estimates of parameters after a few observations are likely to be described by a normal distribution. Second, while a single observation W^n may be non-normal, if we have to use repeated observations (in a single trial) to overcome the nonconcavity of information, then the combined effect of multiple observations is likely to be accurately described by a normal distribution.

4.12 WHY DOES IT WORK?*

4.12.1 Derivation of the knowledge gradient formula

It is not necessary to know how to derive the knowledge gradient formula in order to be able to use it effectively, but sometimes it is nice to go past the "trust me" formulas and see the actual derivation. The presentation here is more advanced (hence the * in the section title), but it is intended to be tutorial in nature, with additional steps that would normally be excluded from a traditional journal article.

The knowledge gradient method is characterized by simplicity and ease of use. In every time step, we can compute $\nu_x^{KG,n}$ for each alternative x by plugging the current values of $\bar{\mu}_x^n$ and $\bar{\sigma}_x^{2,n}$ into the formulas (4.10) and (4.11), and then applying (4.13). After that, our experimental decision is given by $X^{KG,n} = \arg\max_x \nu_x^{KG,n}$, the alternative with the largest knowledge gradient value.

However, it is worthwhile to go through the derivation of the knowledge gradient formula at least once. Not only does this make the KG formulas look less unwieldy, by showing how they originate from the definition of the knowledge gradient, it also gives a sense of how we might go about creating a knowledge gradient method in other optimal learning problems. Later on in this chapter, we derive KG formulas for ranking and selection problems that use some of the non-Gaussian learning models from Chapter 2. This can be done using the same approach as for the independent normal case, but the resulting KG formulas will be very different from the formulas for the normal-normal model.

The expression in (4.4) gives a generic definition of a KG policy, in terms of the expected improvement made by running an experiment. We can write down this expectation in many different settings, but the way we compute it (if, indeed, we can compute it at all) will vary from problem to problem. Often, deriving a computable form for the KG policy poses a computational challenge in research. Thus, the KG approach is a very general idea, but the algorithmic realization of that idea is heavily problem-specific.

We now show how $\nu_x^{KG,n}$ is derived for a particular choice of alternative x at time n. At time n, the estimates $\bar{\mu}_x^n$ and $\bar{\sigma}_x^{2,n}$ are known to us for all x. However, the future estimates $\bar{\mu}_x^{n+1}$ are still random, because we have not yet decided on the $(n+1)$st experiment. It is important to remember that, because the problem is sequential, each new time step changes what is random and what is known. For example, the nth estimate $\bar{\mu}_x^n$ is a random variable from the point of view of any time $n' < n$, but it is a constant from the point of view of time n, when the first n observations have been irrevocably made.

Our goal is to compute (4.3), which requires us to find

$$\mathbb{E}\left[V^{n+1}\left(S^n\right) \,\middle|\, S^n, x^n = x\right] = \mathbb{E}\left[\max_{x'} \bar{\mu}_{x'}^{n+1} \,|\, S^n, x^n = x\right].$$

We assume that we measure x at time n, and examine how this experiment will affect our beliefs about the best alternative. Fortunately, we can simplify our expression for $\max_{x'} \bar{\mu}_{x'}^{n+1}$. Recall from the updating equations (3.2) and (3.3) that $\bar{\mu}_{x'}^{n+1} = \bar{\mu}_{x'}^{n}$ for any $x' \neq x^n$. Thus, we can rewrite

$$\mathbb{E}\left[\max_{x'} \bar{\mu}_{x'}^{n+1} \,\middle|\, S^n, x^n = x\right] = \mathbb{E}\left[\max\left(\max_{x' \neq x} \bar{\mu}_{x'}^{n}, \bar{\mu}_{x}^{n+1}\right) \,\middle|\, S^n, x^n = x\right]. \qquad (4.41)$$

From the point of view of time n, the quantity $\max_{x'} \bar{\mu}_{x'}^{n+1}$ is merely a maximum of a single random variable and a constant. It is typically much easier to compute the expected value of such a quantity than the maximum of multiple random variables.

Next, we consider the conditional distribution of $\bar{\mu}_{x}^{n+1}$ given S^n and $x^n = x$. From the updating equations, we know that

$$\bar{\mu}_{x}^{n+1} = \frac{\beta_x^n}{\beta_x^n + \beta_x^W} \bar{\mu}_x^n + \frac{\beta_x^W}{\beta_x^n + \beta_x^W} W_x^{n+1}, \qquad (4.42)$$

a weighted average of a constant $\bar{\mu}_x^n$ and a random variable W_x^{n+1}. Given our beliefs at time n, the conditional distribution of the true value μ_x of x is normal with mean $\bar{\mu}_x^n$ and variance $\bar{\sigma}_x^{2,n}$. Then, given μ_x, the observation W_x^{n+1} is itself conditionally normal with mean μ_x and variance σ_x^2. What we need, however, is the conditional distribution of W_x^{n+1} given our beliefs at time n, but not given the true value μ_x, and we can find this distribution by computing the moment-generating function

$$\begin{aligned}
\mathbb{E}\left(e^{-rW_x^{n+1}}\right) &= \mathbb{E}\left(\mathbb{E}\left(e^{-rW_x^{n+1}} \,\middle|\, \mu_x\right)\right) \\
&= \mathbb{E}\left(e^{-r\mu_x} e^{\frac{1}{2}\sigma_x^2 r^2}\right) \\
&= e^{\frac{1}{2}\sigma_x^2 r^2} \mathbb{E}\left(e^{-r\mu_x}\right) \\
&= e^{\frac{1}{2}\sigma_x^2 r^2} e^{-r\bar{\mu}_x^n} e^{\frac{1}{2}\bar{\sigma}_x^{2,n} r^2} \\
&= e^{-r\bar{\mu}_x^n} e^{\frac{1}{2}\left(\sigma_x^2 + \bar{\sigma}_x^{2,n}\right) r^2}.
\end{aligned}$$

This is clearly the moment-generating function of the normal distribution with mean $\bar{\mu}_x^n$ and variance $\sigma_x^2 + \bar{\sigma}_x^{2,n}$. The variance can also be written in precision notation as $1/\beta_x^W + 1/\beta_x^n$.

It follows that the conditional distribution of $\bar{\mu}_x^{n+1}$ given S^n and $x^n = x$ is also normal, since $\bar{\mu}_x^{n+1}$ is a linear function of W_x^{n+1} by (4.42). We can find the mean and variance of this distribution by computing

$$\begin{aligned}
\mathbb{E}\left[\bar{\mu}_x^{n+1} \,|\, S^n, x^n = x\right] &= \frac{\beta_x^n}{\beta_x^n + \beta_x^W} \bar{\mu}_x^n + \frac{\beta_x^W}{\beta_x^n + \beta_x^W} \mathbb{E}\left[W_x^{n+1} \,|\, S^n, x^n = x\right] \\
&= \frac{\beta_x^n}{\beta_x^n + \beta_x^W} \bar{\mu}_x^n + \frac{\beta_x^W}{\beta_x^n + \beta_x^W} \bar{\mu}_x^n \\
&= \bar{\mu}_x^n
\end{aligned}$$

and

$$\begin{aligned}
Var\left[\bar{\mu}_x^{n+1}\,|\,S^n, x^n = x\right] &= \left(\frac{\beta_x^W}{\beta_x^n + \beta_x^W}\right)^2 Var\left[W_x^{n+1}\,|\,S^n, x^n = x\right] \\
&= \left(\frac{\beta_x^W}{\beta_x^n + \beta_x^W}\right)^2 \left(\frac{1}{\beta_x^W} + \frac{1}{\beta_x^n}\right) \\
&= \frac{\beta_x^W}{\beta_x^n\left(\beta_x^W + \beta_x^n\right)}.
\end{aligned}$$

Using the definition of the precision again, we can write

$$\begin{aligned}
\frac{\beta_x^W}{\beta_x^n\left(\beta_x^W + \beta_x^n\right)} &= \frac{\bar{\sigma}_x^{2,n}}{\sigma_x^2\left(\frac{1}{\sigma_x^2} + \frac{1}{\bar{\sigma}_x^{2,n}}\right)} \\
&= \frac{\bar{\sigma}_x^{2,n}}{1 + \sigma_x^2/\bar{\sigma}_x^{2,n}}
\end{aligned}$$

which is precisely $\tilde{\sigma}_x^{2,n}$ by (4.9).

We have found that the conditional distribution of $\bar{\mu}_x^{n+1}$, given S^n and $x^n = x$, is normal with mean $\bar{\mu}_x^n$ and variance $\tilde{\sigma}_x^{2,n}$. In words, when we measure x at time n, we expect that the next observation will be equal to $\bar{\mu}_x^n$, that is, our beliefs $\bar{\mu}_x^n$ are accurate on average. As a consequence of this experiment, the variance of our beliefs about μ_x will decrease by an amount equal to the variance of the next observation.

Now, we can return to (4.41) and rewrite the right-hand side as

$$\mathbb{E}\left[\max\left(\max_{x'\neq x}\bar{\mu}_{x'}^n, \bar{\mu}_x^{n+1}\right)\,\middle|\, S^n, x^n = x\right] = \mathbb{E}\left[\max\left(\max_{x'\neq x}\bar{\mu}_{x'}^n, \bar{\mu}_x^n + \tilde{\sigma}_x^n Z\right)\right]$$

where Z is a standard normal (mean 0, variance 1) random variable. Our goal is now to compute an expectation of a function of Z, which looks far more tractable than the expression we started out with.

Let $C_x^n = \max_{x'\neq x}\bar{\mu}_{x'}^n$, and observe that

$$\max\left(C_x^n, \bar{\mu}_x^n + \tilde{\sigma}_x^n Z\right) = \begin{cases} C_x^n & \text{if } Z \leq \frac{C - \bar{\mu}_x^n}{\tilde{\sigma}_x^n} \\ \bar{\mu}_x^n + \tilde{\sigma}_x^n & \text{otherwise.}\end{cases}$$

Thus,

$$\begin{aligned}
\mathbb{E}\left[\max\left(C_x^n, \bar{\mu}_x^n + \tilde{\sigma}_x^n Z\right)\right] &= \int_{-\infty}^{\frac{C_x^n - \bar{\mu}_x^n}{\tilde{\sigma}_x^n}} C_x^n \phi(z)\,dz \\
&\quad + \int_{\frac{C_x^n - \bar{\mu}_x^n}{\tilde{\sigma}_x^n}}^{\infty} \left(\bar{\mu}_x^n + \tilde{\sigma}_x^n z\right)\phi(z)\,dz
\end{aligned}$$

where

$$\int_{-\infty}^{\frac{C_x^n - \bar{\mu}_x^n}{\tilde{\sigma}_x^n}} C_x^n \phi(z)\,dz = C_x^n \Phi\left(\frac{C_x^n - \bar{\mu}_x^n}{\tilde{\sigma}_x^n}\right)$$

and

$$\int_{\frac{C_x^n-\bar{\mu}_x^n}{\tilde{\sigma}_x^n}}^{\infty}\left(\bar{\mu}_x^n+\tilde{\sigma}_x^n z\right)\phi\left(z\right)dz=\bar{\mu}_x^n\Phi\left(-\frac{C_x^n-\bar{\mu}_x^n}{\tilde{\sigma}_x^n}\right)+\tilde{\sigma}_x^n\phi\left(\frac{C_x^n-\bar{\mu}_x^n}{\tilde{\sigma}_x^n}\right).$$

The first term in the second equation uses the symmetry of the normal distribution, and the second term is due to the fact that

$$\int_y^{\infty} z\phi\left(z\right)dz=\phi\left(y\right),$$

which can be verified by a back-of-the-envelope calculation.

Observe now that, due to the symmetry of the normal density,

$$\phi\left(\frac{C_x^n-\bar{\mu}_x^n}{\tilde{\sigma}_x^n}\right)=\phi\left(-\frac{C_x^n-\bar{\mu}_x^n}{\tilde{\sigma}_x^n}\right)=\phi\left(\zeta_x^n\right),$$

which gives us one of the terms that make up the KG formula. To obtain the other term, we consider two cases. If $C_x^n\leq\bar{\mu}_x^n$, then $\bar{\mu}_x^n=\max_{x'}\bar{\mu}_{x'}^n$ and

$$\begin{aligned}C_x^n\Phi\left(\frac{C_x^n-\bar{\mu}_x^n}{\tilde{\sigma}_x^n}\right)+\bar{\mu}_x^n\Phi\left(-\frac{C_x^n-\bar{\mu}_x^n}{\tilde{\sigma}_x^n}\right) &= C_x^n\Phi\left(\zeta_x^n\right)+\bar{\mu}_x^n-\bar{\mu}_x^n\Phi\left(\zeta_x^n\right)\\ &= \bar{\mu}_x^n+\tilde{\sigma}_x^n\left(\frac{C_x^n-\bar{\mu}_x^n}{\tilde{\sigma}_x^n}\right)\Phi\left(\zeta_x^n\right)\\ &= \bar{\mu}_x^n+\tilde{\sigma}_x^n\zeta_x^n\Phi\left(\zeta_x^n\right).\end{aligned}$$

However, if $\bar{\mu}_x^n<C_x^n$, then $C_x^n=\max_{x'}\bar{\mu}_{x'}^n$ and

$$\begin{aligned}C_x^n\Phi\left(\frac{C_x^n-\bar{\mu}_x^n}{\tilde{\sigma}_x^n}\right)+\bar{\mu}_x^n\Phi\left(-\frac{C_x^n-\bar{\mu}_x^n}{\tilde{\sigma}_x^n}\right) &= C_x^n-C_x^n\Phi\left(\zeta_x^n\right)+\bar{\mu}_x^n\Phi\left(\zeta^n\right)x)\\ &= C_x^n+\tilde{\sigma}_x^n\zeta_x^n\Phi\left(\zeta_x^n\right).\end{aligned}$$

Either way,

$$C_x^n\Phi\left(\frac{C_x^n-\bar{\mu}_x^n}{\tilde{\sigma}_x^n}\right)+\bar{\mu}_x^n\Phi\left(-\frac{C_x^n-\bar{\mu}_x^n}{\tilde{\sigma}_x^n}\right)=\max_{x'}\bar{\mu}_{x'}^n+\tilde{\sigma}_x^n\zeta_x^n\Phi\left(\zeta_x^n\right).$$

Putting all the terms together,

$$\begin{aligned}\mathbb{E}\left[\max\left(C_x^n,\bar{\mu}_x^n+\tilde{\sigma}_x^n Z\right)\right] &= \max_{x'}\bar{\mu}_{x'}^n+\tilde{\sigma}_x^n\left(\zeta_x^n\Phi\left(\zeta_x^n\right)+\phi\left(\zeta_x^n\right)\right)\\ &= \max_{x'}\bar{\mu}_{x'}^n+\tilde{\sigma}_x^n f\left(\zeta_x^n\right),\end{aligned}$$

whence

$$\begin{aligned}\nu_x^{KG,n} &= \mathbb{E}\left[\max_{x'}\bar{\mu}_{x'}^{n+1}-\max_{x'}\bar{\mu}_{x'}^n\;\middle|\;S^n,x^n=x\right]\\ &= \mathbb{E}\left[\max\left(\max_{x'\neq x}\bar{\mu}_{x'}^n,\bar{\mu}_x^{n+1}\right)\;\middle|\;S^n,x^n=x\right]-\max_{x'}\bar{\mu}_{x'}^n\\ &= \tilde{\sigma}_x^n f\left(\zeta_x^n\right),\end{aligned}$$

which is precisely the KG formula that we introduced earlier.

The above derivation relies on our assumption of independent alternatives, as well as our use of a normal-normal learning model. However, in the process, we followed a set of steps that can be used to derive knowledge gradient formulas for other learning problems, as well. In particular, we use the exact same approach to derive KG formulas for ranking and selection problems with non-Gaussian learning models later in this chapter. The steps are:

1) We calculate the conditional distribution, given S^n and $x^n = x$, of the value function $V^{n+1}\left(S^{n+1}(x)\right)$. In ranking and selection with independent alternatives, this reduces to the problem of finding the conditional distribution of $\bar{\mu}_x^{n+1}$ (in the above derivation, this was normal). Recall that this is called the predictive distribution of $\bar{\mu}_x^{n+1}$, because it is what we predict about time $n+1$ at time n.

2) We calculate the conditional expectation of $V^{n+1}\left(S^{n+1}(x)\right)$ over the predictive distribution. This is especially simple in ranking and selection with independent alternatives, because it is simply an expected value of a function of a one-dimensional random variable (in the above derivation, this was the standard normal random variable Z).

3) We subtract the quantity $V^n(S^n)$, which is viewed as deterministic at time n.

In most of the problems discussed in this book, the real challenge lies in the second step. For example, if we introduce correlations into the problem, the predictive distribution of the vector $\bar{\mu}^{n+1}$ is still fairly simple, but it is a bit more difficult to calculate the expectation of the value function $V^{n+1}\left(S^{n+1}(x)\right)$ over this distribution. However, there are many learning problems that seem complicated, but that actually yield simple closed-form expressions for the knowledge gradient. Later on, we encounter many more varied examples of such problems.

4.13 BIBLIOGRAPHIC NOTES

Section 4.2 - The idea of collecting information based on the expected value of a single experiment was first introduced by Gupta & Miescke (1994) and Gupta & Miescke (1996). The concept of the knowledge gradient was developed in greater depth in Frazier et al. (2008). We connect the KG concept to the idea of information economics from Chapter **??**; in fact, Chick & Gans (2009) describes a KG-like approach as the "economic approach to simulation selection."

Section 4.3 - The material on the non-concavity of information is based on Frazier & Powell (2010). Additional discussion on the nonconcavity of information can be found in Weibull et al. (2007), which also presents conditions where it is not always optimal to choose the alternative that appears to be best.

Section 4.5 - The knowledge gradient for correlated beliefs was first introduced in Frazier et al. (2009).

Section 4.7 - The development of the knowledge gradient for non-Gaussian distributions in this section is mostly new. A version of KG for the gamma-exponential model was presented in Ryzhov & Powell (2011*c*).

Section **??** - The expected improvement algorithm was proposed by Jones et al. (1998*a*) for problems with continuous decisions. This algorithm is also sometimes known as efficient global optimization or EGO in this setting. Some more recent work on EI can be found in Gramacy & Lee (2011). The $LL(N)$ methodology was originally put forth by Chick & Inoue (2001), with extensive empirical validation undertaken by Inoue et al. (1999) and Branke et al. (2005). A general overview of the LL approach is available in Chick (2006). The LL_1 procedure is a more recent development laid out in Chick et al. (2010).

PROBLEMS

4.1 Your estimate of the long-run performance of a mutual fund was that it returns 8 percent, but your distribution of belief around this number is normally distributed with a standard deviation of 3 percent. You update this each year (assume that successive years are independent), and from history you estimate that the standard deviation of the return in a particular year is 6 percent. At the end of the most recent year, the mutual fund returned -2 percent.

a) Use Bayesian updating to update the mean and standard deviation.

b) What do we mean when we say that the "normal distribution is conjugate"?

c) What will be the precision of my estimate of the long-run performance in four years (after four experiments, starting with the current belief state)?

4.2 You have three places that serve takeout food around your area and you want to maximize the quality of the total food you intake over time. The three restaurants are:

1. Vine Garden (VG)
2. Wise Sushi (WS)
3. Food Village (FV)

You assume a normal prior with (μ_x, β_x) on the quality of the food in these places and the experiments are normally distributed with precision $\beta^W = 1$.

1. Define the expected opportunity cost (EOC) for this setting (assume that the discount factor, $\gamma = 1$).
2. For a single ω, Table 4.5 below contains your prior, the truth and the outcomes of your observations until the third time step $n = 3$. Write down the empirical opportunity cost after the third observation (note that you will need to update your priors).
3. If you are valuing your current utility from food higher than future time periods $(\gamma < 1)$, how would you expect the behavior of the optimal policy to change as opposed to having $\gamma = 1$?

4.3 Table 4.8 shows the priors $\bar{\mu}^n$ and the standard deviations σ^n for five alternatives.

Iteration	VG	WS	FV
Prior (μ_x, β_x)	$(6,2)$	$(7,1)$	$(8,1)$
Truth for μ_x	5	8	7
1	4		
2		9	
3			8.5

Table 4.5 Priors and observations

a) Compute the knowledge gradient for each alternative in a spreadsheet. Create a plot with the mean, standard deviation and the knowledge gradient for each alternative.

b) Three of the alternatives have the same standard deviation, but with increasing priors. Three have the same prior, but with increasing standard deviations. From these two (overlapping) sets of alternatives, describe how the knowledge gradient changes as we vary priors and the standard deviation of our belief.

Choice	$\bar{\mu}^n$	σ^n
1	3.0	8.0
2	4.0	8.0
3	5.0	8.0
4	5.0	9.0
5	5.0	10.0

Table 4.6 Calculations illustrating the knowledge gradient index

4.4 Assume that we have a standard normal prior about a true parameter μ which we assume is normally distributed with mean $\bar{\mu}^0$ and variance $(\sigma^0)^2$.

a) Given the observations $W^1, \ldots, W^n$, is $\bar{\mu}^n$ deterministic or random?

b) Given the observations $W^1, \ldots, W^n$, what is $\mathbb{E}(\mu|W^1, \ldots, W^n)$ (where μ is our truth)? Why is μ random given the first n experiments?

c) Given the observations $W^1, \ldots, W^n$, what is the mean and variance of $\bar{\mu}^{n+1}$? Why is $\bar{\mu}^{n+1}$ random?

4.5 As a venture capitalist specializing in energy technologies, you have to decide to invest in one of three strategies for converting solar power to electricity. A major concern is the efficiency of a solar panel, which tends to run around 11 to 12 percent. You are at the point where you are running field experiments, but each field experiment produces an estimate of the efficiency which has a standard deviation of 4.0. The first technology appears to have an efficiency of 11.5 percent, but the standard deviation in your distribution of belief around this number is 2.0. The second technology has an estimated efficiency of 11.0 with a standard deviation of 3.5, while the third has an estimated efficiency of 12.0, with a standard deviation of 1.5. You want to choose the technology with the highest efficiency.

a) Use the knowledge gradient to tell you which technology you should experiment with next.

b) If you are only going to do one last experiment, is this the optimal choice? Explain.

c) If you did not make this last investment, you would choose technology 3 with an efficiency of 12.0. What is the expected efficiency of the technology that would be chosen as best after the last investment?

4.6 Consider the problem of finding the best person to serve as the lead-off hitter on a baseball team. The lead-off hitter is evaluated primarily for his ability to get on base. If x is the hitter, his outcome would be recorded as a Bernoulli random variable W^n_x, where $W^n_x = 1$ if he gets on base, and $W^n_x = 0$ otherwise. We are going to conduct these experiments during spring training, where we are primarily focused on finding the best lead-off hitter (we do not really care about his performance while we are collecting the information). Learning is accomplished using the beta-Bernoulli model. Each alternative x has an unknown success probability ρ_x, and our goal is to find the alternative with the highest success probability. We begin with a prior belief $\rho_x \sim Beta\left(\alpha^0_x, \beta^0_x\right)$. Supposing that we measure alternative x^n at time n, our beliefs are updated according to the equations

$$\alpha^{n+1}_x = \begin{cases} \alpha^n_x + W^{n+1}_x & \text{if } x^n = x \\ \alpha^n_x & \text{otherwise,} \end{cases}$$

$$\beta^{n+1}_x = \begin{cases} \beta^n_x + \left(1 - W^{n+1}_x\right) & \text{if } x^n = x \\ \beta^n_x & \text{otherwise,} \end{cases}$$

where the observation W^{n+1} is equal to 1 with probability ρ_x and 0 with probability $1 - \rho_x$. Recall that, under this model, our estimate of the uncertain truth ρ_x at time n is $\mathbb{E}\left(\rho_x \,|\, S^n\right) = \frac{\alpha^n_x}{\alpha^n_x + \beta^n_x}$.

a) Suppose that we measure alternative x at time n. Show (by conditioning on the truth) that

$$\mathbb{P}\left(W^{n+1}_x = 1 \,|\, S^n\right) = \frac{\alpha^n_x}{\alpha^n_x + \beta^n_x}, \qquad \mathbb{P}\left(W^{n+1}_x = 0 \,|\, S^n\right) = \frac{\beta^n_x}{\alpha^n_x + \beta^n_x}.$$

b) Use the definition of the knowledge gradient to write out an expression for the knowledge gradient for this problem. You do not have to reduce the expression in any way (for example, you will have an expectation, but you do not have to reduce it to a convenient expression).

4.7 Garrett Jones was a minor leaguer in baseball trying to break into the major leagues. He was called up to play in a few major league games, where he made one hit in eight at bats. After this weak performance, he was sent back to the minor leagues. The major league club that was evaluating him is looking for someone who can hit at a certain level against an existing major league hitter. Think of this as choosing between an uncertain minor leaguer, and a more certain major leaguer (so this is a case with two alternatives).

a) It is reasonable to assume that no-one would ever make a decision based on a single at bat. Assume that our minor leaguer will be given at least 10 at bats, and that we will now assume that our prior belief about his batting average is normally distributed with mean 0.250 and standard deviation 0.20. Further assume that our belief about the major leaguer is also normally distributed with mean 0.267 and standard deviation of 0.10. Finally assume that we are going to approximate the observed batting average from at least 10 at bats as normally distributed with mean:

$$W_{minor} = \frac{H}{m}$$

where H is the number of hits and m is the number of at bats. The variance of W_{minor} is given by

$$\sigma^2_W = \rho_{minor}(1 - \rho_{minor})/m$$

where $\rho_{minor} = .235$ is the expected batting average of the minor leaguer (the true batting average is a random variable). Give the expression for the knowledge gradient resulting from m at bats and compute the knowledge gradient.

b) Assume that the knowledge gradient for a single at bat is very small. Without actually computing the knowledge gradient, plot what is likely the general shape of the value of observing m at bats as a function of m. If you were going to use the KG(*) policy, what would you do? What are the implications of this shape in terms of how a coach should evaluate different minor leaguers?

4.8 Consider a ranking and selection problem with exponential observations and gamma priors. That is, if we choose to measure alternative x at time n, we observe $W^{n+1}_x \sim Exp\left(\lambda_x\right)$. The rate λ_x is unknown, but we start with the assumption that $\lambda_x \sim Gamma\left(\alpha^0_x, \beta^0_x\right)$ and update these beliefs as we run experiments. Our beliefs about the alternatives are independent. Thus, if we measure alternative x at time n, we update $\alpha^{n+1}_x = \alpha^n_x + 1$ and $\beta^{n+1}_x = \beta^n_x + W^{n+1}_x$, while keeping $\alpha^{n+1}_y = \alpha^n_y$ and $\beta^{n+1}_y = \beta^n_y$ for all $y \neq x$.

a) Suppose that our objective is to find the largest rate λ_x. Define the knowledge gradient of alternative x at time n as

$$\nu^{KG,n}_x = \mathbb{E}\left[\max_y \frac{\alpha^{n+1}_y}{\beta^{n+1}_y} - \max_y \frac{\alpha^n_y}{\beta^n_y} \,|\, S^n, x^n = x\right].$$

Argue that

$$\mathbb{E}\left[\max_y \frac{\alpha^{n+1}_y}{\beta^{n+1}_y} \,|\, S^n, x^n = x\right] = \mathbb{E}\left[\max\left(C^n_x, \frac{\alpha^n_x + 1}{Y}\right)\right]$$

where $C_x^n = \max_{y \neq x} \frac{\alpha_y^n}{\beta_y^n}$ and $Y \sim Pareto\,(\alpha_x^n, \beta_x^n)$.

(Remember that the $Pareto\,(a,b)$ density is $g\,(t) = \frac{ab^a}{t^{a+1}}$ for $t > b$ and zero elsewhere.)

b) Suppose that $\frac{\alpha_x^n+1}{\beta_x^n} \leq C_x^n$. Show that

$$\mathbb{E}\left[\max\left(C_x^n, \frac{\alpha_x^n+1}{Y}\right)\right] = C_x^n.$$

c) Suppose that $\frac{\alpha_x^n+1}{\beta_x^n} > C_x^n$. Show that

$$\mathbb{E}\left[\max\left(C_x^n, \frac{\alpha_x^n+1}{Y}\right)\right] = \frac{\alpha_x^n}{\beta_x^n} + \frac{(\beta_x^n)^{\alpha_x^n}\,(C_x^n)^{\alpha_x^n+1}}{(\alpha_x^n+1)^{\alpha_x^n+1}}.$$

d) Based on parts b) and c), show that

$$\nu_x^{KG,n} = \begin{cases} \frac{(\beta_x^n)^{\alpha_x^n}(C_x^n)^{\alpha_x^n+1}}{(\alpha_x^n+1)^{\alpha_x^n+1}} & \text{if } x = \arg\max_y \frac{\alpha_y^n}{\beta_y^n} \\ \frac{(\beta_x^n)^{\alpha_x^n}(C_x^n)^{\alpha_x^n+1}}{(\alpha_x^n+1)^{\alpha_x^n+1}} - \left(\max_y \frac{\alpha_y^n}{\beta_y^n} - \frac{\alpha_x^n}{\beta_x^n}\right) & \text{if } x \neq \arg\max_y \frac{\alpha_y^n}{\beta_y^n}, \\ & \frac{\alpha_x^n+1}{\beta_x^n} > \max_y \frac{\alpha_y^n}{\beta_y^n} \\ 0 & \text{otherwise.} \end{cases}$$

4.9 We again consider a ranking and selection problem with exponential observations and gamma priors. There are five alternatives, and the belief state for a certain time step n is given in table 4.7. The objective is to find the largest rate.

x	α_x^n	β_x^n
1	2	18
2	3	17
3	1	7
4	2	15
5	3	14

Table 4.7 Priors for exercise 4.9

4.10 Consider a ranking and selection problem with independent alternatives, exponential observations and gamma priors.

a) Suppose that we want to find the alternative with the highest rate. Derive the KG formula given in (4.29). (Hint: Consider the three cases given in the formula separately before trying to take integrals.)

b) Repeat your analysis for the case where our goal is to find the lowest rate. Show that the KG formula is given by (4.30).

4.11 Table 4.8 shows the priors $\bar{\mu}^n$ and the standard deviations σ^n for five alternatives.

a) Compute the knowledge gradient for each alternative in a spreadsheet. Create a plot with the mean, standard deviation and the knowledge gradient for each alternative.

b) Three of the alternatives have the same standard deviation, but with increasing priors. Three have the same prior, but with increasing standard deviations. From these two (overlapping) sets of alternatives, describe how the knowledge gradient changes as we vary priors and the standard deviation of our belief.

Choice	$\bar{\mu}^n$	σ^n
1	3.0	8.0
2	4.0	8.0
3	5.0	8.0
4	5.0	9.0
5	5.0	10.0

Table 4.8 Calculations illustrating the knowledge gradient index

a) Compute $\nu_x^{KG,n}$ for all x. Which alternative will the KG policy measure at time n?

b) Suppose now that $\alpha_2^n = 1$ and $\beta_2^n = 6$, while our beliefs about the other alternatives remain unchanged. How does this change your answer to part a)? Why did KG change its decision, even though the estimate $\frac{\alpha_2^n}{\beta_2^n}$ is actually smaller now than what it was in part a)?

4.12 Table 4.9 shows the priors $\bar{\mu}^n$ and the standard deviations σ^n for five alternatives.

a) Compute the knowledge gradient for each in a spreadsheet.

b) You should observe that the knowledge gradients are fairly small. Provide a plain English explanation for why this would be the case.

Choice	$\bar{\mu}^n$	σ^n
1	3.0	4.0
2	4.0	6.0
3	20.0	3.0
4	5.0	5.0
5	6.0	7.0

Table 4.9 Priors for exercise 4.12

4.13 In Section 4.12.1, we showed that $\mathbb{E}\left[\bar{\mu}_x^{n+1} \,|\, S^n, x^n = x\right] = \bar{\mu}_x^n$ in the normal-normal model. Verify that this also holds for our estimates of the unknown parameters in other learning models:

a) Show that $\mathbb{E}\left[\frac{\alpha_x^{n+1}}{\beta_x^{n+1}}\,\middle|\, S^n, x^n = x\right] = \frac{\alpha_x^n}{\beta_x^n}$ in the gamma-exponential model.

b) Repeat part a) for the gamma-Poisson model.

c) Show that $\mathbb{E}\left[\frac{\alpha_x^{n+1}}{\alpha_x^{n+1}-1} b_x^{n+1}\,\middle|\, S^n, x^n = x\right] = \frac{\alpha_x^n}{\alpha_x^n-1} b_x^n$ in the Pareto-uniform model.

d) Show that $\mathbb{E}\left[\frac{\alpha_x^{n+1}}{\alpha_x^{n+1}+\beta_x^{n+1}}\,\middle|\, S^n, x^n = x\right] = \frac{\alpha_x^n}{\alpha_x^n+\beta_x^n}$ in the beta-Bernoulli model.

In each of these cases, our estimates of the unknown parameters (the rate λ_x, the upper endpoint B_x, and the success probability ρ_x) are expected to stay the same on average. That is, given that we are at time n, we expect that the estimate will not change on average between time n and time $n + 1$. This is called the *martingale property*.

4.14 Consider a ranking and selection problem with independent alternatives, Poisson observations and gamma priors. Suppose that the objective is to find the alternative with the highest Poisson rate. Show that the KG formula is given by (4.31).

4.15 Consider a ranking and selection problem with independent alternatives, uniform observations and Pareto priors. Suppose that the objective is to find the largest upper endpoint among the uniform distributions.

a) Show that the predictive distribution of b_x^{n+1} given S^n and $x^n = x$ is a mixed discrete/continuous distribution given by

$$\mathbb{P}\left(b_x^{n+1} = b_x^n \mid S^n, x^n = x\right) = \frac{\alpha_x^n}{\alpha_x^n + 1}$$

and

$$f\left(y \mid S^n, x^n = x\right) = \frac{1}{\alpha_x^n + 1} \frac{\alpha_x^n \left(b_x^n\right)^{\alpha_x^n}}{y^{\alpha_x^n + 1}}, \qquad y > b_x^n.$$

b) Show that the KG formula for alternative x in this problem is given by (4.32).

4.16 Consider a ranking and selection problem with independent alternatives, Bernoulli observations and beta priors. Suppose that the objective is to find the alternative with the largest success probability. Show that the KG formula is given by (4.26), and verify that this formula remains the same (aside from changing the maximum in the definition of C_x^n to a minimum) if we change the objective to finding the lowest success probability instead of the highest.

4.17 Consider a ranking and selection problem with independent alternatives, normal observations and normal priors. However, instead of the usual objective function $F^\pi = \max_x \bar{\mu}_x^n$, we use

$$F^\pi = \sum_x \mathbb{E}\left[\left(\mu_x - \bar{\mu}_x^n\right)^2 \,\middle|\, S^N\right].$$

That is to say, instead of trying to find the alternative with the highest value, our goal is now to run experiments in such a way as to reduce (on average) the sum of squared errors of our final estimates at time N of all the values. Derive the KG formula for this problem.

4.18 Suppose that we have four different products. The profit margin of product x, $x = 1, 2, 3, 4$ is represented by μ_x. We can choose to perform a market study on product x to get an observation W_x of its profit margin. The observation is normally distributed with mean μ_x and variance $(\sigma_x^W)^2$. Table 4.10 below gives the true values of μ_x and $(\sigma_x^W)^2$:

x	μ_x	$\left(\sigma_x^W\right)^2$
1	15	4
2	10	3
3	12	5
4	11	2

Table 4.10 Priors for exercise 4.18.

However, we do not know the true values of μ_x. We describe our beliefs about them using a multivariate Gaussian prior with the following mean vector and covariance matrix:

$$\bar{\mu}^0 = \begin{bmatrix} 12 \\ 14 \\ 13 \\ 10 \end{bmatrix}, \qquad \Sigma^0 = \begin{bmatrix} 12 & 0 & 6 & 3 \\ 0 & 7 & 4 & 2 \\ 6 & 4 & 9 & 0 \\ 3 & 2 & 0 & 8 \end{bmatrix}.$$

Assume that we choose to observe product 3 and we observe $W_3^1 = 15$. Show how our beliefs would change using the updating equations

$$\bar{\mu}^1 = \bar{\mu}^0 + \frac{W_x^1 - \bar{\mu}_x^0}{\left(\sigma_x^W\right)^2 + \Sigma_{xx}^0} \Sigma^0 e_x, \qquad \Sigma^1 = \Sigma^0 - \frac{\Sigma^0 e_x e_x^T \Sigma^0}{\left(\sigma_x^W\right)^2 + \Sigma_{xx}^n} \tag{4.43}$$

where $x = 3$ is the particular product that you are considering, and e_x is a column vector of zeroes with a single 1 in the xth coordinate. Report the resulting values of $\bar{\mu}^1$ and Σ^1. (Equation (4.43) gives the "convenient" version of the updating equations, where you don't have to compute an inverse. You do not have to derive these equations.)

4.19 Consider a ranking and selection problem with independent alternatives, normal observations and normal priors. However, instead of the usual objective function $F^\pi = \max_x \bar{\mu}_x^n$, we use

$$F^\pi = \max_x |\bar{\mu}_x^n| .$$

That is, we want to find the alternative with the largest absolute value. Show that the KG factor of alternative x is given by

$$\nu_x^{KG,n} = \tilde{\sigma}_x^n \left(f\left(\zeta_x^n\right) + f\left(\delta_x^n\right) \right),$$

where

$$\zeta_x^n = -\left| \frac{\bar{\mu}_x^n - \max_{x' \neq x} |\bar{\mu}_{x'}^n|}{\tilde{\sigma}_x^n} \right|, \qquad \delta_x^n = -\frac{\bar{\mu}_x^n + \max_{x' \neq x} |\bar{\mu}_{x'}^n|}{\tilde{\sigma}_x^n.}$$

4.20 The revenue generated by an online advertisement has an exponential distribution with parameter λ (thus the mean revenue is $\frac{1}{\lambda}$). We do not know λ, so we use a gamma prior and assume $\lambda \sim Gamma(a, b)$ for $a > 1$. Recall that the density of the gamma distribution is given by

$$f(x) = \frac{b(bx)^{a-1}e^{-bx}}{\Gamma(a)},$$

where $\Gamma(a) = (a-1)!$ if a is integer. The mean of $f(x)$ is a/b and the variance is a/b^2.

a) What is the current belief about the value of λ? That is, if you had to guess the value of λ, what would you say, and why? If you assume that λ is exactly equal to this belief, what is the mean revenue generated by the advertisement?

b) Now take an expectation of the mean revenue over the entire distribution of belief. That is, compute $\mathbb{E}(\frac{1}{\lambda})$ for $\lambda \sim Gamma(a, b)$.

c) Why are your answers to (a) and (b) different? Which one should you actually use as your estimate of the mean reward, and why?

4.21 Consider a ranking and selection problem with normal observations and normal priors (and independent beliefs).

a) Create a MATLAB file called `kg.m` which implements the KG policy. As a template, you can use the code that was first introduced in exercise 3.2 which can be downloaded from

```
http://optimallearning.princeton.edu/exercises/exploration.m
http://optimallearning.princeton.edu/exercises/explorationRun.m
```

b) Set $N = 5000$, $M = 50$ and report the confidence interval.

c) Set $N = 1$, $M = 1$ and run the policy 100 times. How often does KG find the best alternative?

4.22 You would like to find the price of a product that maximizes revenue. Unknown to you, the demand for the product is given by

$$D(p) = 100e^{-.02p}.$$

Total revenue is given by $R(p) = pD(p)$. Assume prices are integers between 1 and 100.

You set a price and watch it for a week. Assume that the observed revenue R^n in week n is given by

$$R^n = R(p) + \epsilon^n$$

where $\epsilon^n \sim \mathcal{N}(0, 400^2)$. However, since you believe that the function is continuous, you realize that your beliefs are correlated. Assume that your belief about $R(p)$ and $R(p')$ is correlated with covariance function

$$Cov^0(R(p), R(p')) = 400^2 e^{-0.03|p-p'|}.$$

So, $Cov^0(R(p), R(p)) = Var^0(R(p)) = 400^2$, and $Cov^0(R(20), R(30)) = 400^2 \times 0.7408$. Use this to create your prior covariance matrix Σ^0. Assume that your initial estimate of $R(p)$ is $\bar{\mu}^0_p = 2000$ for each p (this is known as a "uniform prior," and represents a situation where you have no idea which price is the best.) Note that we are using online learning for this exercise.

a) Write out the updating formulas for updating your estimate $\bar{\mu}^n_p$ giving the estimated revenue when you charge price p, and the updating formula for the covariance matrix Σ^n.

b) Implement the algorithm for computing the knowledge gradient in the presence of correlated beliefs (call this algorithm KGCB), using as a starting point

`http://optimallearning.princeton.edu/exercises/KGCorrBeliefs.m`

An example illustration of the KGCB algorithm is given in

`http://optimallearning.princeton.edu/exercises/KGCorrBeliefsEx.m`

Verify your algorithm first by running it with a diagonal covariance matrix, and showing that the independent and correlated KG algorithms give you the same numbers. Note that you may find that for some prices, the knowledge gradient is too small to compute (for example, you get a very negative exponent).

c) Next use the initial covariance matrix described above. Plot the log of the knowledge gradient for prices between 1 and 100 (again, be careful with large negative exponents), and compare your results to the log of the knowledge gradient assuming independent beliefs. How do they compare?

d) Now we are going to compare policies. Please do the following:

   i) Run your KGCB algorithm for 10 experiments, and plot after each experiment the opportunity cost, which means you take what you think is the best price based on your current set of estimates, and compare it to the revenue you would get if you knew the best price (hint: it is \$50). Repeat this exercise 20 times, and report the average opportunity cost, averaged over the 20 iterations. The goal here is to get a sense of the variability of the performance of a learning policy.

   ii) We want to compare KGCB against pure exploration, pure exploitation and interval estimation using $z_\alpha = 1.96$ (a standard default). For each policy, perform 20 sample paths and plot the average opportunity cost over all 20 sample paths. Compare all three policies. [Please make sure that you are resetting the covariance matrix to its initial structure after you are done with a sample path (once you have gone over a single truth). If you skip this part, the KG algorithm will assume it has very precise priors for the second run, which of course is not true.]

e) The knowledge gradient policy can be quite sensitive to the prior. Instead of an initial prior of 2000, now assume that we start with a uniform prior of 500 (same standard deviation). If you edit the prior (column B), the package regenerates the truth. You are going to have to re-enter the formula for the truth after changing the prior. After doing this, perform 3 repetitions of KG, pure exploration and pure exploitation, and contrast their performance.