## Optimizing an Outcome

If you are using your model to control or optimize a situation then presumably there are one or more parameters which you can control; you'll want to choose those parameters to optimize the outcome.

To do this effectively you need to have a good understanding of two things:

- which parameters you can control and
- what exactly it is that you want to achieve. In other words, how exactly would you decide whether or not one set of parameters is better than another?

**Example:** You are modeling the traffic through an intersection in order to determine the timing of the traffic lights that will optimize the traffic flow.

Parameters over which we presumably have no control include the rates at which cars arrive along each of the incoming roads.

Parameters that we can control include the timings of the lights.

How will we decide when one timing is better than another? In other words, what does it mean for the traffic flow to be better in one situation than it is in another? Possibilities include:

- the longest amount of any time any one individual has to wait for the light is less,
- the average length of time that individuals have to wait for the light is less,
- the average length of time that it takes a car to get through the intersection is less,
- the length of the longest line of traffic at the light is less etc.

It may be good to look at one or more of these measures; it may be that minimizing one of these measures is equivalent to minimizing another or you may want to minimize a (possibly weighted) average of two or more of these measures.

Having determined the parameters over which you have control and the quantity that you want to optimize you then have the following situation:

- Let $\Omega$ denote the space of all the parameters which you can control and
- let $\phi : \Omega \to \mathbb{R}$ denote the function that is to be maximized or minimized.

The problem is to find $\omega \in \Omega$ that optimizes $\phi$. This is an optimization problem just like the one that we discussed above for determining the values of the parameters given data. We can use the same techniques as there; analysis if the function is simple enough (usually it isn't) or numerical optimization techniques when it isn't so simple.

## Sensitivity analysis

You should have a sense now of how hard it can be to determine the values of the parameters. This means that we often have poor estimates of their values, so it's important to understand what happens to the quantities of interest when you change the parameters.

This is called a *sensitivity analysis*.

One way to determine the sensitivity of a quantity to the value of a parameter is by simulation; you take each parameter (typically one at a time) and change its value by, say, 1% and see what the corresponding change is in the value of the quantity.

Small (percentage) changes in the value of the quantity are good; this means that even if your parameters are a little off, the quantity of interest won't change too much.

This leads us to a definition of the sensitivity of a quantity $Q$ to a parameter $p$.

The sensitivity of a quantity $Q$ to a parameter $p$ tells us the percentage amount that $Q$ will change when $p$ is changed by a small percentage.

To get an analytical expression for it, let $\Delta p$ denote a small change in the value of $p$ and let $\Delta Q$ denote the corresponding change in $Q$ (when the other parameters are held constant). Since $\Delta Q/\Delta p \approx \partial Q/\partial p$ we have

$$\frac{\Delta Q/Q}{\Delta p/p} \approx \left(\frac{\partial Q}{\partial p}\right)\left(\frac{p}{Q}\right)$$

This is the *sensitivity* of $Q$ to $p$. We denote it $S(Q,p)$, so

$$\frac{\Delta Q}{Q} \approx S(Q,p)\frac{\Delta p}{p}$$

The following example is a simple, clean example where we are able to perform a sensitivity analysis analytically.

**Example:**

A counter is used to measure the radioactive decay in a sample of radioactive material. Each radioactive decay that is detected by the counter locks the counter for $3 \times 10^{-9}$ seconds, during which time any decays that occur aren't measured. How can we determine the half-life of the material from the measurements made by the counter?

The variables of interest in this problem are:

- The times between radioactive decays $T$ and the times between decays that are measured by the counter $S$.
- The number of radioactive atoms in the sample $X$.

The parameters are:

- The half-life of the substance. This is a quantity that we want to determine.
- The lock time $a = 3 \times 10^{-9}$ is a parameter whose value is probably only approximately known and could vary from counter to counter.
- The initial number of atoms in the substance $N$.

We will assume the following:

- There is only one type of radioactive material in the sample and it doesn't have a complicated sequence of decays; the isotope decays to another isotope that doesn't then decay further.
- The decays occur "at random" independent of when they have occurred in the past.
- The over all rate of decay is proportional to the number of radioactive atoms in the sample.

These assumptions mean that $T$ is a random variable; it takes on different values apparently at random.

From observations of $S$ we want to estimate the time it takes for half of the material to decay.

Radioactive decay is one of the classical examples of exponential decay and the very concept of half-life is based on this model. However, how can we relate this to observations of $S$?

According to quantum theory, it is impossible to predict when an atom will decay. However, the tendency of an atom to decay is constant over time.

In other words, the probability that the atom will decay in a small interval of time $\Delta t$ is proportional to the length of the interval and this doesn't change over time. Let's call the constant of proportionality $p$.

This sounds like a description of a Poisson process except that since there is only one atom there can be at most one event.

However, the atoms in a sample of material decay independently of each other. This means that the decays that occur in a sample occur as a Poisson process with a rate $\lambda$ that is proportional to the number of atoms in the sample; $\lambda = pX$.

Notice that every time an atom decays, $X$ decreases by 1, so the underlying rate of decay keeps changing.

However, initially $X = N$ which is very large (on the order of $10^{23}$) so any change in $X$ that is less than a million or so is quite negligible.

So we'll approximate and say that the rate is constant and equal to $pN$ atoms per second.

Thus, the random variable $T$ is the time between events in a Poisson process, so $T$ is exponentially distributed with an average of $1/pN$ seconds where $p$ is a number that depends on the material that is decaying.

Because of the lock time, the value of $S$ is not necessarily the same as the value of $T$; sometimes it is equal to the value of $T$ but other times it is equal to the sum of two or more successive values of $T$.

We can think of $S$ as being composed of two parts, the lock time equal to $3 \times 10^{-9}$ plus the time starting from the end of the lock time until the next decay $R$:

$$S = 3 \times 10^{-9} + R.$$

Since the events in a Poisson process occur at times that are independent of when they have occurred in the past, this means that whenever we look at the process (be it immediately after an event or at some point in between events) the time we have to wait until the next event will still have an exponential distribution with the same parameter $1/pN$.

This means that the distribution of $R$ is the same as the distribution of $T$! Thus

$$E(S) = 3 \times 10^{-9} + E(R) = 3 \times 10^{-9} + E(T) = 3 \times 10^{-9} + \frac{1}{pN}.$$

In other words

$$p = \frac{1}{(E(S) - 3 \times 10^{-9})N}.$$

To get the half-life of the material from $p$ we need to think about the exponential decay model of radioactive decay and how that relates to our model.

In our model, the number of atoms in the material is a random variable, so the time it takes for the number to decrease to half its original value is *random*.

However, the exponential decay model is obtained as the expected decay of $X$ as follows.

In a small interval of time $\Delta t$ the expected number of decays is equal to $pX\Delta t$. In other words

$$E(\Delta X) = -pX\Delta t.$$

Assuming the variable decays according to its expected value (this is reasonable if we look at the material for a lengthy period of time because those intervals where it is greater than its expected value and those where it is less than its expected value will, with high probability, roughly cancel out) it follows that

$$\frac{\Delta X}{\Delta t} = -pX.$$

Treating $X$ as a continuous variable (it is so large that it is reasonable to do this) and taking a limit as $\Delta t$ goes to 0 we get

$$X' = -pX$$

which means that $X$ decays exponentially with rate $p$. Thus, the half-life of the material is $\tau = (\ln 2)/p$ which can be written in terms of $S$ as

$$\tau = \left(E(S) - 3 \times 10^{-9}\right) N \ln 2.$$

To estimate the half-life from observations of $S$ recall that by the strong law of large numbers we know that the average of a large number of successive values of $S$ is approximately equal to $E(S)$, so

$$\tau \approx \left( \frac{s_1 + s_2 + \ldots + s_n}{n} - 3 \times 10^{-9} \right) N \ln 2$$

where $s_1, s_2, \ldots, s_n$ are the observed values of $S$.

Our estimate of $\tau$,

$$\tau_e = \left( \frac{s_1 + s_2 + \ldots + s_n}{n} - a \right) N \ln 2$$

contains the parameters $N$ and the lock time $a = 3 \times 10^{-9}$. Both of these are probably only roughly known.

How sensitive is $\tau_e$ to the values of these parameters?

The sensitivity of our estimate to the initial number of atoms in the sample $N$ is

$$S(\tau_e, N) = \left(\frac{\partial \tau_e}{\partial N}\right)\left(\frac{N}{\tau_e}\right) = 1$$

In other words, everything else being equal, a percentage error in the value of $N$ will produce an equal percentage error in the estimation of $\tau_e$.

On the other hand, the sensitivity of $\tau_e$ to $a$ is

$$S(\tau,_e a) = \left(\frac{\partial \tau_e}{\partial a}\right)\left(\frac{a}{\tau_e}\right) = \frac{(N \ln 2)a}{\left(\frac{s_1+s_2+\ldots+s_n}{n} - a\right) N \ln 2} = \frac{a}{\left(\frac{s_1+s_2+\ldots+s_n}{n} - a\right)}$$

The value of this depends on the data and on the value of $a$.

Our estimate of $a$ is $3 \times 10^{-9}$ and suppose, for example, that the average of the $s_i$'s is equal to $6 \times 10^{-9}$ then $S(\tau, a) = 0.5$ telling us that, everything else being equal, a percentage error in the value of $a$ will produce half that percentage error in the value of $\tau_e$.

Notice, moreover, that the value of $S(\tau_e, a)$ depends on $\bar{s} = (1/n)\sum_i s_i$ which is an estimate for $E(S)$. In particular, the larger $\bar{s}$ is, the less sensitive $\tau_e$ is to errors in $a$. But $E(S)$ is large when the sample has fewer decays per second, so we can decrease our sensitivity to errors in $a$ by reducing the size of the sample.

Another source of error may be in the measurements of the $s_i$'s. How accurately does the counter measure these values? Notice that if the number of observations $n$ is large, then we can expect the errors in these measurements to cancel out over the many observations, so the over-all error will probably decrease the larger $n$ is.

Even if the $s_i$'s were measured exactly, the derivation of the formula for $\tau_e$ replaced $E(S)$ with the average of the $s_i$'s, $\bar{s}$. We wouldn't expect $\bar{s}$ to be exactly equal to $E(S)$, but the larger $n$ is the closer it probably is.

The study of these two sources of error belongs to the theory of estimation in statistics. We will touch on this theory very briefly here.

Let's look first at what effect measurement error might have on our estimate of $\tau$.

Let's suppose that $s_1, s_2, \ldots s_n$ were the actual values of $S$ that occurred but the values that we *measured* were $\tilde{s}_1, \tilde{s}_2, \ldots \tilde{s}_n$. Denote the errors by $\varepsilon_i = s_i - \tilde{s}_i$.

We will use our actual measurements when we estimate $\tau$ of course, so our estimate of $\tau$ will be

$$\left( \frac{\tilde{s}_1 + \tilde{s}_2 + \ldots + \tilde{s}_n}{n} - 3 \times 10^{-9} \right) N \ln 2$$

when it should have been

$$\left( \frac{s_1 + s_2 + \ldots + s_n}{n} - 3 \times 10^{-9} \right) N \ln 2.$$

The error is

$$\left( \frac{\varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_n}{n} \right) N \ln 2.$$

To understand how large the error is likely to be we need to understand its probability distribution.

Measurement errors are generally at least approximately normally distributed, presumably with mean equal to 0 and some standard deviation that we'll call $\sigma$.

It is a property of normal distributions that when you add two variables that are both normally distributed, their sum is also normally distributed.

Moreover, the errors from one measurement to the next are probably independent. This means that the sum $\varepsilon_1 + \varepsilon_2 + \ldots \varepsilon_n$ is normally distributed with mean equal to 0 and standard deviation equal to $\sqrt{n}\sigma$, so the error above is normally distributed with mean equal to 0 and standard deviation equal to $\frac{N\sigma \ln 2}{\sqrt{n}}$.

Since the standard deviation decreases with $n$ we can make this error as small as we want (with high probability) by taking many measurements. Moreover, if we know $\sigma$ then, for a fixed value of $n$, we can put a bound on the size that the error is likely to be.

Now let's look at the error that we get by replacing $E(S)$ with $\bar{s}$.

We know by the strong law of large numbers that if $n$ is large then the value of $\bar{s}$ is, with high probability, close to $E(S)$. But how close is close and how many measurements do we need?

Notice that the value of $\bar{s}$ is itself a random variable; if we looked at another $n$ measurements of $S$ the values of the $s_i$'s would be different, so the value of $\bar{s}$ would be different.

This means that $\bar{s}$ has a probability distribution; some values of $\bar{s}$ are more likely than others.

The Central Limit Theorem tells us that if $x_1, x_2, \ldots x_n$ are independent (or weakly dependent) observations of a random variable $X$ then the distribution of $\bar{x} = (x_1 + x_2 + \ldots + x_n)/n$ looks more and more normal the larger $n$ is.

This is true *whatever distribution $X$ has* (practically).

Moreover, our formulae above for the mean and standard deviation of a sum of independent variables tells us that the mean of $\bar{x}$ is $E(X)$ and the standard deviation of $\bar{x}$ is $\mathrm{SD}(X)/\sqrt{n}$.

To return to the example, recall that $S = 3 \times 10^{-9} + R$ where $R$ is exponentially distributed and the underlying Poisson process has rate $pN$.

This means that $R$ has mean equal to $1/pN$ and standard deviation equal to $1/pN$ (since the standard deviation of an exponential distribution is equal to its mean).

It follows that the mean of $S$ is $3 \times 10^{-9} + 1/pN$ and the standard deviation is $1/pN$.

Thus, if $n$ is large, the distribution of $\bar{s}$ is approximately normal with mean equal to $E(S)$ and standard deviation equal to $1/pN\sqrt{n}$.

Since almost all values of a variable lie within 3 standard deviations of the mean, this means that $\bar{s}$ is highly likely to lie within $E(S) \pm 3/pN\sqrt{n}$.

Thus, when we use $\bar{s}$ instead of $E(S)$ to estimate $\tau$ the error that we are likely to incur has a maximum size of $3 \ln 2/p\sqrt{n}$. We can make this error small by taking many observations of $S$.