Protocol on making practical estimates in cases of large uncertainty

Owen Cotton-Barratt

Version 0.3, January 2013

1 Introduction

Making reasonable estimates from limited data, as in Fermi problems, is widely accepted as a useful method of quickly finding orders of magnitude, or of performing sanity checks on work. In many contexts this is the limit of such techniques, and to make decisions one should find more precise data. This is sensible when such data is available, but sometimes it is necessary to make decisions in the absence of good data. In such cases it is useful to be able to make estimates, and to be able to minimise and quantify the level of uncertainty.

The purpose of these notes is to provide some guidance on how to produce and record estimates in situations where there are a lot of unknowns, and how to use these estimates.

There are some general issues which accompany radical uncertainty, and measures which can be taken to account for these. They cannot always be separated from the case at hand, but to the extent to which it is possible it is helpful to do the analysis once and give people more time and brainspace to spend on the specifics of the individual scenario.

It is hoped that by providing a standard protocol we will make different estimates more comparable, since there will be fewer worries that, for example, different authors mean different things by 'lower bound'. Even if there are systematic biases in the protocol (which we strive to avoid), so long as it is applied consistently the values it produces should be more comparable than estimates produced by differing methods with differing systematic biases. We also provide advice on how to collect and keep data, both to more easily allow updating with the addition of information, and to make it as forward-compatible as we can for new methods or standards of producing estimates from the data.

I would like to thank Nick Beckstead, Nic Dunkley, Toby Ord, and Rob Wiblin for useful and insightful comments.

2 Modelling Uncertainty

Most forms of decision theory tell you to use expectations in cases of uncertainty. But to take an expectation, you need a belief about the distribution the values are drawn from. We will look at practical approximations we might use to produce such an estimated distribution.

Fermi calculations usually arrive at the quantity to be determined by multiplying together (or dividing by) several other, easier to estimate, values. There will be some uncertainty associated with each of these. When we consider the product, the uncertainty will combine in some way. We are going to use a relatively straightforward model of how this combination occurs.

The central limit theorem and its relations let us know that if enough independent random variables are summed, the result approximates a normal distribution. Equivalently, if enough (positive) independent random variables are multiplied, the result converges on a log-normal distribution.[1] While some components of our calculations involve addition, cost effectiveness analysis most frequently features multiplication and division.

When we have a quantity y that we want to estimate, we will try to decompose it as a product $y = \prod_{i=1\ldots n} x_i$ of positive values x_i, where we model each x_i as drawn from a random variable X_i expressing our uncertainty of the true value. Consequently y is assumed to be drawn from the random variable $Y = \prod_{i=1\ldots n} X_i$. We would ideally do this so that the individual X_i are independent. Sometimes this may be impossible, but if we can estimate the degree of correlation we can try to correct for the correlation. Were still investigating the best methods for such corrections.

In the case where the X_i are all independent, we will estimate the mean and variance of $\log(X_i)$ for each X_i. Summing these will give the mean and variance of $\log(Y)$, and the assumption that Y is a log-normal distribution will complete the model. Useful data about our overall beliefs about Y (as captured by this model) can then be read off, such as its expectation, median, and 95% confidence levels.

2.1 Warnings

It is important to be aware that this model is not always appropriate.

2.1.1 Strict positiveness

A common obstruction is the requirement that each X_i be positive.[2] If this is not naturally the case, there are a couple of approximations we might use.

If it is almost certain that X_i is positive (and not large and negative), and a small probability of zero or negative value is unimportant, we might simply model it as positive even though we are not sure. Otherwise we might look to

[1] Note that we will use natural logarithms throughout.
[2] Or really that it be of known sign; a factor of -1 is easy to deal with.

express it as a difference $X_i = X_i^+ - X_i^-$ of two positive random variables; in this case Y will be a difference of two log-normal distributions.

2.1.2 Correlation

If some of the X_i are correlated, this will not change the mean of the normal distribution $\log(Y)$, but it will affect the variance. Collecting estimates for the amount of pairwise correlation could be done for example by producing estimates for the distribution X_j while assuming that x_i takes a certain value, and seeing how the estimates of X_j vary with the assumed values for x_i.

More complex correlations are unfortunately harder to capture (the pairwise correlation data is not enough). We would advise trying to find a decomposition of y into factors which display no more than pairwise correlation (of any significance; treating small correlations as zero is probably reasonable). This may involve aggregating sets of the x_i into single quantities to be estimated; estimating bounds for these aggregates may be harder but should take correlation into account. Alternatively it may involve decomposing some x_i into their constituent components.

There is an additional complication. Even when we have no particular expectation towards correlation or anticorrelation, this is not the same as independence. It is often quite plausible that some single factor is throwing off our estimates of two quantities, making them either correlated or anticorrelated. While this should have little effect on the median of the eventual distribution, it seems it should make the tails fatter. Effectively this would increase the variance in our end model. It is likely that we should have a corrective term for this effect, but it is not clear what the best form of such an effect would be.

This area is in need of more research, and this section should eventually have more details added to it. For the moment we somewhat ignore questions of correlation in the sequel.

3 Taking estimates from one person

We will now give a protocol for recording estimates from a single person.

First, decompose y as above as a product, and check that the person is happy that uncertainty in the factors is more or less uncorrelated. If this isn't the case, combine or separate terms until the correlation only occurs between two factors at a time.

3.1 Procedure for a single value

We are now in the situation of trying to estimate the distribution of X_i. Sometimes we will have good reasons to expect X_i to be of a certain distribution, and then we can read off the mean and variance of $\log(X_i)$. In the absence of such a belief, we will use the default assumption that X_i is itself distributed as a log-normal; it is often the case that the x_i could be further decomposed as a product of other variables.

We are going to describe how to collect estimated values and use these to derive the parameters of the log-normal distribution we use to model X_i. There may also be cases where there are reasons to expect X_i to have a different type of distribution, but where we don't know the parameters. In that case a similar procedure should let you produce parameters to use. A future version of these notes may detail the type of situation where we would expect particular distribution types, and precise instructions for calculating parameters for each, but for now we restrict to the log-normal case.

It should be straightforward at this point, but because humans are prone to err and display systematic biases, there are some questions about how best to proceed.

Question 1 *What are the best questions to ask people to elicit useful and accurate responses?*

Ideally we would answer this empirically by testing with surveys.[3]

For the time being, I suggest we first ask first for the median estimate m (the fiftieth percentile). Then we should produce a range R around m and ask for the estimated probability that the true value lies in R. If possible ask someone else to produce the range R. Else as a default consider $[m/2, 2m]$, but be prepared to adjust this if it is unappropriate for the case in hand. We might also present people with the distribution inferred from their answers and give them an opportunity to adjust it.

Remark 2 *Note that more than for any of the other unanswered questions, it would pay to get a good system in early. If we later have improvements to the calculation procedure, it's easy to go back and correct earlier work just by changing the spreadsheet. But recollecting data will be time consuming or impossible.*

Now we will use these estimates to provide our estimates for the mean and variance of $\log(X_i)$. Again, the best method of doing this might be determined empirically:

Question 3 *What is the most accurate procedure for estimating mean and variance of* $\log(X_i)$ *from provided estimates?*

If we have a model for how people produce their estimates, we can use a regression here; this could be sensible if it is automated. Giving people an opportunity to see the distribution produced and adjust may serve a similar

[3] In a blog post, Robin Hanson wrote: Weve also learned better ways to elicit estimates. For example, instead of asking for a 90% confidence interval on a number, it is better to ask for an interval, and then for a probability. It works even better to ask about an interval someone else picked. Also, instead of asking people directly for their confidence, it is better to ask them how much their opinion would change if they knew what others know. However these results have not yet been published. Source: http://www.overcomingbias.com/2012/11/wanted-elite-crowds.html

role, but requires a more active role on the part of the estimators, and is not amenable to rapidly changing the method later on.

Remarks 4 *It is possible that we should multiply the variance by some constant to try to counteract the overconfidence bias, but for the moment I will not do this. Ideally we might test people with similar questions where the answers are known, to calibrate for their overconfidence.*

3.1.1 Probabilities

One fairly common kind of parameter we might try to estimate is a probability. While it is fair to assume it is above zero, it is clearly unreasonable to assume a log-normal distribution, since probabilities cannot be larger than one.

The standard technique used by statisticians for probabilities seems to be to assume a Beta distribution.

Remark 5 *I'd like to have a better explanation of why this is best. Logit-normal seems intuitively appealing to me, but is a hassle to work with as it has things like a non-analytic mean.*

3.1.2 Dates

This is not quite relevant, since dates rarely appear directly in Fermi calculations. However they are often used in questions that figure out how precise or calibrated peoples confidence estimates are. They are also used in forecasting when future events will occur. In this case, in order to model them as a log-normal distribution it is probably usually best to set it up as a question of how long ago an event occurred before the present, or how long in the future an event will occur. In some cases it might be appropriate to anchor on a different date (for example if one event is known to have followed another you might estimate the time between them).

4 Taking estimates from multiple people

By taking estimates from more than one person, you can hope to minimise the effects of biases or just outlying estimates from a single person. This can be done by the straightforward method of asking a lot of people, or for some kinds of data perhaps through other means such as prediction markets.

For the time being we give notes on things to bear in mind when collecting and collating data, rather than a definitive procedure; while we may later add more precise guidelines, we recognise that this stage will necessarily vary with the time and resources available to investigate the question.

4.1 Choosing who to ask

There are a couple of conflicting pressures here. On the one hand it is important to ask people who will think clearly about the matter and take relevant factors into account. On the other, it can be best to avoid taking estimates from a group who have been talking to each other a lot about the matter, as groupthink may obstruct good independent estimates. You will also want to ask people who have a good enough understanding of the situation to be doing more than pure guessing.

In any case, it is likely best to start by taking factorisation of y into independent x_i; it is not necessary to ask the same collection of people for estimates of each x_i.

4.1.1 With unlimited resources

Ideally, you might proceed as follows. Produce a summary sheet of the major considerations and arguments which might affect our beliefs about x_i, even ones you think are invalid (but of course, include the rebuttals). It is likely to be best to avoid discussing numbers here (except factually certain ones), to avoid anchoring effects on estimates.

Separately, give this sheet to several intelligent people (ideally with a good track record of estimating) who have not previously thought much about the questions.[4] Then get them to give estimates.

For certain questions, where there are potentially a lot of important considerations for accurately giving the answer, and we do not have any reason to believe we have found them all, it may be better to avoid giving summary notes at all, for fear of inducing a groupthink-like phenomenon. It is likely you would only want to even consider doing this in situations where the estimators are all quite knowledgable about the issue.

Remark 6 *I'm not sure about this last point; could that really be better than trying to pool all thoughts of the crucial considerations and then follow the above procedure? Need to think more about this, or perhaps someone else knows.*

4.2 How to take an average

Remark 7 *This is likely another area where a good method can be empirically determined. Perhaps there is already research on this?*

The straightforward approach would be to use as our estimate for the mean of $\log(X_i)$ the mean of the estimates computed from each person's answers. For

[4] This last restriction might seem a bit perverse. We are trying to avoid a certain kind of selection bias, where the people who have sought out and thought about particular questions are likely on average to overestimate the importance of those questions, and perhaps values related to their answers. It is hard to say how large the effects of this sort of bias may be.

the variance we would do something similar, but perhaps also consider a term which looks at the variance of the distribution of estimates of the mean.[5]

That approach, though, is based on the idea that the estimates themselves are more-or-less drawn from the distribution of uncertainty around the value. In practice it is often the case that the most extreme estimates in either direction are the result of some misconception or calculational error, and it can be better to discard these altogether to avoid them distorting the average. The median is more robust to these extreme errors than the mean.

Remark 8 *I asked whether there was a standard statistical technique for this, and apparently there is. The idea is to have a model for how everyone's estimates are drawn, and run a regression to get an idea of the most likely underlying values. It is better if you have some calibration data for the people, since some people may be systematically better estimators than others. This looks like it is probably the best way to approach this, but wants more software support to make it practical.*

5 Using estimates

The end result of the estimation process is a model of $\log(Y)$ as a normal distribution with mean μ and standard deviation σ. We'll briefly discuss some reasons you might use different quantities from this output.

5.1 Expectation

The expectation (or mean) of Y is a value that is often called for in decision theory if y captures something you care about. For the log-normal distribution, the expectation is $e^{\mu+\sigma^2/2}$; note that this is above the median and for cases of high variance may be well above, because so much of the expected value lies in the positive tail.

Sometimes the value you care about may be $1/y$; since $\log(1/Y)$ is a normal distribution with mean $-\mu$ and standard deviation σ, the expectation of $1/Y$ is $e^{\sigma^2/2-\mu}$.

5.2 Median

The median value of Y is simply e^μ. You might use this if considering questions asking which of two quantities is likely to be higher.

5.3 Confidence bounds

People are often interested in what levels we can confidently say y lies above. This is easily read off from the normal distribution. For example, Y should lie

[5] I'm not certain how much to weight this, and it's worth being aware that it will be likely to be a systematic underestimate, because what it captures is the variance between different people's estimates of the answer, and these may all be skewed by relying on the same data.

above $e^{\mu-1.65\sigma}$ around 95% of the time, and above $e^{\mu-2.33\sigma}$ around 99% of the time.

These figures can be useful in circumstances where you want to be careful not to overstate your case, although they will often be far lower than our best guess for the true value.

There may be some other reasons to err on the side of caution. In situations where using this style of estimation to see which of several quantities has the highest expectation, we have to deal with a selection bias: it may often be the case that the frontrunner leads just because it got lucky and our estimates were out in the positive direction. This particularly runs the risk of producing over-optimistic results for those projects we are most uncertain about.

5.4 Directing future research

Seeing which factors contribute the most uncertainty in an estimate gives easy guidance for where further research to try to give more precise estimates would be valuable, or where it would be essentially irrelevant.

In the case mentioned above, where looking for which quantity has the highest expectation, it is particularly easy to focus research efforts. It will become most important to find ways to reduce the uncertainty among the most promising looking options among current estimates, especially those which have greatest uncertainty. It is important to be aware that further research will typically push values back towards the mean (or our prior probability)! However when it does increase expectation it can do so by a large amount.