R-G Theory Propagation of Error
George Rebane
30 January 2021

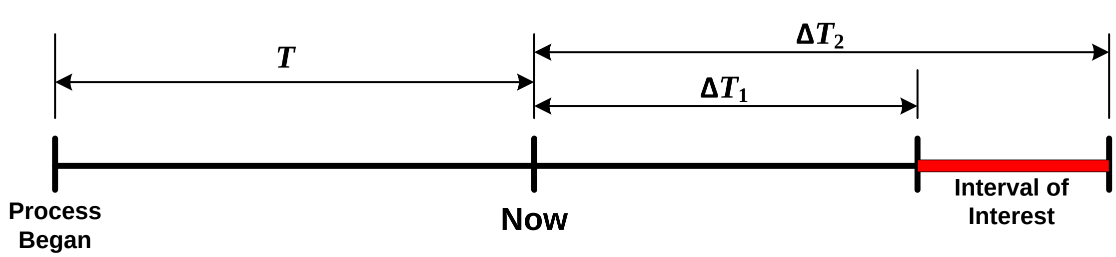
Readers were first introduced to my extension of Richard Gott's work (here) in developing a means to estimate the remaining lifetime of the human race to within a certain probability. This work turned out to have greater importance and utility as an extended theory on the probabilistic duration of minimally known processes (MKPs). In technical fields and even everyday life MKPs abound, and their ending in or survival beyond a future time interval of interest would provide useful information for making many kinds of decisions. This motivated the development of the R-G theory of MKPs.

In the following I give a brief summary of the sources of error in the calculation of probabilities for the cases when only the age of the MKP is known, and when only the MKP's lifetime is known. The complete derivation of the error propagation formulas is found in *TN2011-2: R-G Error Propagation*. However, before diving into that, we detour a bit to review the import of MKPs and where they fit into our everyday lives.

First, any ongoing thing over time can be considered as a 'process'. This includes anything from the longtime periodic eruption of Old Faithful, a price trend for a stock, or the duration of a period without rain (drought?), or bicycle race on a circuit over city streets, or when someone at work will quit their job, or when the next meteor will streak across the sky, or when someone at a DMV window will finish their business, or … . You get the idea. The only requirement to properly calculate R-G probabilities is that the observed process fits the 'minimally known' requirement.

For MKPs of unknown duration or lifetime, all you know is how long the process has been going on or the age of the process when you encountered or began observing it. Processes for which only their lifetime or duration is known, and which you begin observing at some unknown time during this interval also qualify as MKPs. For both kinds you can use R-G to compute the probability that the process ends or not within a specified future time interval.

Let's take a look at each of these and do a couple of examples. First, MKPs of a known age T look like the one shown below.



Note that the time interval of interest does not have to start right now; it can be of arbitrary length ($\Delta T_2 - \Delta T_1$) starting at any time ΔT_1 in the future. The probability P that the process ends some time during the interval of interest is given by

$$P = \frac{T\left(\Delta T_2 - \Delta T_1\right)}{\left(T + \Delta T_1\right)\left(T + \Delta T_2\right)} \ .$$

It's clear that ΔT_2 is always greater than ΔT_1, and that if we want the interval of interest to start now, we simply let $\Delta T_1 = 0$, which makes the above reduce to the elegantly simple formula

$$P = \frac{\Delta T}{T + \Delta T} \ .$$

As an example of how to use these formulas, suppose you've had your eye on a particular residential house in a nice neighborhood that you would like to buy some time in the near term. All you know is that the current owners have lived there for the last ten years ($T = 10$), and you'd like to know the probability that they'll put that house up for sale some time in the next six months, or $\Delta T = 0.5$. From the above formula we calculate $P = 0.5/(10+0.5) = 0.048$, or just shy of one chance out of 20. That's slim pickings, but you now also know that the current owners will still be living there six months from now with probability $1 – P = 0.954$, which may allow you time to put together the money for an attractive offer that they may accept.

Now suppose that you know that you will come into some funds a year from now which you must invest some time during the following six months for tax reasons. That house is still on your mind, and you'd like to know what the 'chance' is that the owners would still be there and prepared to sell some time during your six months interval of interest. We would now use the first formula above to calculate that probability with $\Delta T_1 = 1$ and $\Delta T_2 = 1.5$. This gives

$$P = \frac{10\left(1.5 - 1\right)}{\left(10 + 1\right)\left(10 + 1.5\right)} = 0.040 \ .$$

As you'd expect, the chance is less that they would sell during any given six-month interval that is farther out in the future. However, the probability that they would still be living in the house 18 months ($\Delta T = 1.5$) from now is $1 – P = 0.870$. Here $P = 0.130$ and is probability, calculated from the second formula, that they will sell the house some time during the next 18 months. And that may still give you an opportunity to come up with another deal to get the house after 18 months from now.

Moving on to the MKP for which you only know its lifetime T_L, the formula that gives the probability of the process terminating during an interval delt, no matter when in the future it occurs, is simply

$$P = \frac{\Delta T}{T_L}$$

So let's go to Yellowstone National Park to view the Old Faithful geyser which erupts regularly on average every 94 or 68 minutes. Your busy park visit schedule has set aside 30 minutes to view Old Faithful. You can go to the geyser now or later after viewing some of the park's other attractions. Since you don't know what the current eruption interval is, you would like to know the chances for seeing Old Faithful during any 30-minute interval whenever you arrive. Well, it's easy to calculate both probabilities as 30/94 = 0.320 and 30/68 = 0.440. This tells you that the least chance you'll have of seeing the geyser when you arrive is about one out of three, and you make your visit plans accordingly.

As another example, let's suppose that you have a friend competing in a closed-circuit bicycle race during which the competitors complete each circuit in about 30 minutes. Your schedule allows you to pick a spot from which to cheer on your friend and stay there only 15 minutes. You plan to arrive after the race has started and want to know the probability that you'll see your friend in the limited time you can observe the race. From the above formula the intuitive answer is also the correct answer – the probability that he will pass your point during any 15-minute observation time interval is simply $P = 15/30 = 0.500$.

Propagation of Errors into the calculation of *P*.

With the above background under our belts, we can now tackle the main event. For both the known-age MKPs and the known-lifetime MKPs we usually know T and T_L only to within certain error limits, or 'ballpark' if you will. The question then is how do errors in those two inputs impact the reliability of the calculated probabilities. The more formal phrasing of this question is – if T and T_L are known only to within a probability distribution (pdf), how does that impact calculation of the desired probabilities, and how reliable will our answers be?

Let's start with the known-age MKP where from its pdf we know T to have its expected value μ_T and standard deviation σ_T. Then from the first two formulas above, we compute the expected termination probabilities as

$$E(P) = \frac{\mu_T(\Delta T_2 - \Delta T_1)}{(\mu_T + \Delta T_1)(\mu_T + \Delta T_2)} + \left[\frac{\Delta T_2}{(\mu_T + \Delta T_2)^3} - \frac{\Delta T_1}{(\mu_T + \Delta T_1)^3}\right]\sigma_T^2 \ ,$$

$$E(P) = \Delta T\left[\frac{(\mu_T + \Delta T)^2 + \sigma_T^2}{(\mu_T + \Delta T)^3}\right]$$

These more complex formulas for the termination probabilities account for the possibility that μ_T does not equal T, the most likely value of age when its uncertainty is not taken into consideration. As we see below, for asymmetric pdfs the expected and most likely values of T differ.

And, as derived in *TN2011-2*, the standard deviations of the above termination probabilities are respectively

$$\sigma_P = \left[\frac{\Delta T_1}{(\mu_T + \Delta T_1)^2} - \frac{\Delta T_2}{(\mu_T + \Delta T_2)^2} \right] \sigma_T \,,$$

$$\sigma_P = \frac{\Delta T}{(\mu_h + \Delta T)^2} \sigma_T$$

Now, for known-lifetime MKPs we again propagate the error in T_L as represented by its pdf with expected value μ_{TL} and standard deviation σ_{TL}. Using the same approach as in the known-age MKP development, in this simpler case the expected termination probability and its standard deviation are shown to be

$$E(P) = \left[1 + \frac{\sigma_{TL}^2}{\mu_{TL}^2} \right] \frac{\Delta T}{\mu_{TL}}, \quad \sigma_P = \left(\frac{\Delta T}{\mu_{TL}} \right) \sigma_{TL} \ .$$

Finally, we look at what kind of probability distributions are available for representing practical uncertainties for the values of age and lifetime. If enough data is available for certain kinds of MKPs to compute such mean values and standard deviations, then we should use these in the above error propagation formulas to compute termination probabilities and estimates of their reliability. However, most MKPs that we run into are usually one or few of-a-kind processes.

For such MKPs we must use our experience or some third-party expertise to develop the needed probability distribution or pdf. And this process itself requires the transforming of experience or sparse relevant information into a subjective framework to represent the needed pdf. Fortunately, there is a convenient and powerful tool for quantifying such subjectively derived parameters. These pdfs are the mode augmented boxcar (MAB) distribution and the extended MAB (EMAB) distribution. These are described in following reports by the present author –

Rebane, G.J., ‘Predicting with Expressed Beliefs – A Formal Approach’ (February 2016)
Rebane, G.J., *TR1509-1: The Mode Augmented Boxcar (MAB) Distribution* (September 2015)
Rebane, G.J., *TR1908-1: The Extended MAB (EMAB) Distribution* (August 2019)