TN2011-1: R-G Probabilities for Discrete Events

George Rebane
5 November 2020 (V18nov20)

References:

1. Rebane, G.J., *TN0708-1: Predicting the Lifetime of Minimally Known Processes – Gott Extended*
2. Rebane, G.J., *TN1411-1: Gott's future duration, ..., and this too shall pass*
3. Rebane, G.J., *TN1902-1: Predicting Termination of Minimally Known Ongoing Processes*

The references document the development of R-G theory for minimally known processes (MKPs) that may terminate at any (continuous) time. The termination probabilities are calculated for the next time interval ΔT from now, and also for some arbitraty ΔT in the future. The last reference [3] extends the theory from knowing only the age T of the MKP to know only its lifetime T_L.

The present development extends the above formulations to MKPs consisting of a sequence of any type of discrete event that occurs with arbitrary time intervals. Here we know only N, the number of times an event E of an ongoing MKP has occurred, and the case where we know only N_L, the total number of times E will occur during the lifetime of an ongoing discrete MKP. Again, the same copernican observation arguments hold as are made in the references.

From [1] we express the probability of termination in the next ΔT as

$$\boxed{P_E = \frac{\Delta T}{T + \Delta T}} \tag{1}$$

For the discrete case we know that N events have occurred at some unknown times during the unknown past interval T. And we argue that a proportionate number n of similar events may be expected to take place in some unknown time interval ΔT starting now. With these arguments we may then state from the analogue of [1] that

$$P(N,n) = \frac{n}{N+n} \tag{2}$$

Where $P(N,n)$ now gives the probability that the process will terminate after completing n_E events $1 \le n_E \le n$. $n_E \neq 0$ by our assumption that the MKP is an ongoing process that has yet to terminate. The probability that the MKP will terminate with the next event – i.e. $n_E = 1$ - is then

$$P(N,1) = \frac{1}{N+1} \tag{3}$$

And as expected

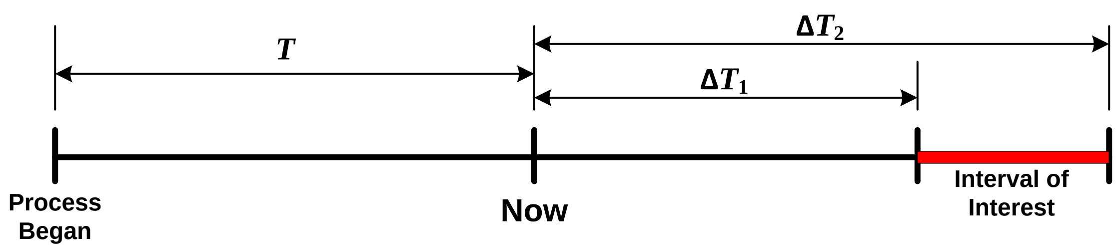
$$\lim_{n\to\infty} P(N,n) = 1 \ , \tag{4}$$

which confirms that the MKP will ultimately terminate with certainty. For this to hold, we may have to place additional restrictions on the MKP by stating that we also know that it is a finite stochastic process with $n_E < \infty$, and may terminate after any number n_E of additional future events.

R-G theory [2] calculates the termination probability during an arbitrary interval that begins at a future time as

$$P(T,\Delta T_1,\Delta T_2) = \frac{T(\Delta T_2 - \Delta T_1)}{(T+\Delta T_1)(T+\Delta T_2)} \ , \tag{5}$$

where the terms are defined in the figure below for $0 \le \Delta T_1 \le \Delta T_2$.



As required, we see that $P(T,\Delta T_1,\Delta T_2) \xrightarrow{\Delta T_1 \to 0} P(T,\Delta T_2) = \dfrac{\Delta T_2}{T+\Delta T_2}$, (6)

as given in (1). With this extension we may calculate the probability of, say, a 5-year-old process terminating in a 3-month window that starts 9 months from now as

$$P(60,9,12) = \frac{60(12-9)}{(60+9)(60+12)} = 0.036 \ . \tag{7}$$

The complement of this probability is that the process will NOT terminate in that interval, i.e. that it will terminate after 60 + 9 + 3 = 72 months. Note that (5) is the conditional probability of terminating during $\Delta T_2 - \Delta T_1$, given that the process did NOT terminate during ΔT_1. This line of reasoning is confirmed by showing that the MKP will terminate at some $t_E > t_{NOW}$. Specifically, that

$$\Pr(t_E < t_1) + \Pr(t_E \in [t_1,t_2]) + \Pr(t_E > t_2) = 1 \tag{8}$$

Where the t_i, i = 1,2 are terminal times of the ΔT_i intervals shown in the above figure. Then substituting from above, we rewrite (8) as

$$P(T,\Delta T_1)+\frac{T(\Delta T_2-\Delta T_1)}{(T+\Delta T_1)(T+\Delta T_2)}+\left[1-P(T,\Delta T_2)\right]=1$$

$$P(T,\Delta T_1)-P(T,\Delta T_2)+\frac{T(\Delta T_2-\Delta T_1)}{(T+\Delta T_1)(T+\Delta T_2)}=0$$

$$\frac{\Delta T_1}{T+\Delta T_1}-\frac{\Delta T_2}{T+\Delta T_2}+\frac{T(\Delta T_2-\Delta T_1)}{(T+\Delta T_1)(T+\Delta T_2)}=0$$

$$-\frac{T(\Delta T_2-\Delta T_1)}{(T+\Delta T_1)(T+\Delta T_2)}+\frac{T(\Delta T_2-\Delta T_1)}{(T+\Delta T_1)(T+\Delta T_2)}=0,\ \text{QED}$$

Now we ask the analog question for discrete MKPs – 'what is the probability that the process will halt within the n events that follow an arbitrary number n_F of future events?'

Letting $n_2 = n_F + n$ and following similar reasoning used to arrive at (5), we have for $n_2 > n_F$

$$P(N,n_F,n)==\frac{N(n_2-n_F)}{(N+n_F)(N+n_2)}=\frac{Nn}{(N+n_F)\left[N+(n_F+n)\right]} \tag{9}$$

Again, we show that (2) is recovered from

$$P(N,n_F,n_2)\xrightarrow{n_F\to 0,\ n_2\to n}P(N,n)=\frac{n}{N+n}$$

If we want the probability that the process will end exactly at $n_E = n_F$ events from now, then in (9) we simply set $n = 1$ and n_F to $n_F - 1$, the number of events preceding the n_E event. This gives

$$P(N,n_E)=P(N,n_F-1,1)=\frac{N}{(N+n_F-1)(N+n_F)} \tag{10}$$

Say, $N = 100$, we desire to know the probability that the process stops exactly after event $n_E = 10$ events from now. Then

$$P(100,10)=\frac{100}{(100+10-1)(N+10)}=0.008 \tag{11}$$

Finally, in the last instance of such MKPs we seek the discrete analog of knowing the lifetime T_L of a process, but not when in that interval we will observe it – the copernican perspective. The continuous version of this case was developed in [2].

Suppose we have a MKP that consists of N_L events which define its lifetime. We do not know how many of those events have already occurred, but want to observe the next n events and know what the probability is of the process terminating with one of the observed events. Using the reasoning introduced in [2] and further developed above, the desired probability is simply

$$P\left(N_L, n\right) = \frac{n}{N_L} \tag{12}$$

And if we desire the probability of the process ending during some future sequence of n events that starts n_F events from now and ends with $n_2 = n_F + n > n_F$ events from now, we reason as in [3].

$$\begin{aligned} P_E\left(N_L, n_F, n\right) &= \Pr\left(\text{MKP survives through } n_F \text{ events}\right)\ldots \\ &\quad \Pr\left(\text{MKP halts within next } n \text{ events} \,|\, \text{survival}\right) \\ &= \left[1 - P\left(N_L, n_F\right)\right] P\left(N_L, n\right) \\ &= \left(1 - \frac{n_F}{N_L}\right)\left(\frac{n}{N_L - n_F}\right) = \left(\frac{N_L - n_F}{N_L}\right)\left(\frac{n}{N_L - n_F}\right) = \frac{n}{N_L} \end{aligned} \tag{13}$$

which is the same as (12) where n is the total number of events of interest in the future sequence. For the lifetime version (vs the age version) of MKPs we need specify only the number n of sequential events of interest, and not where they are located in the remaining sequence of events before the MKP terminates.

As for the continuous case in (8), it is straightforward to show that

$$\Pr\left(n_E \le n_F\right) + \Pr\left(n_E \in \left[n_F + 1, n_F + n\right]\right) + \Pr\left(n_E > n_F + n\right) = 1 \tag{14}$$

Finally, the scalability arguments for continuous time processes made in the references apply equally for the discrete case using scaling constants of the form $h = n/N$.