Implicit Assumptions in the Application of Gott's Formula

Kevin R. Anderson
Institute for Mathematics and Its Applications
University of Minnesota
Minneapolis Minnesota 55455

January 24, 2000

A method has been recently suggested [1] for building confidence intervals for the time remaining in an epoch using for data only the length of time since the beginning of the epoch. This method, here-to-fore referred to as "Gott's Formula", depends on the "Copernican principle in time", and is an extremely clever and powerful idea. Its power lies in the fact that it does not depend on particular aspects of the system being studied, but only on the fact that we cannot be assumed to occupy any particular place in the epoch. This generality has lead to its being applied to a wide variety of systems, from the age of the universe, and the length of human existence [1], to the duration of Broadway plays [2] and the period for which a political party will hold power [3]. In this letter I show that there is a wide class of natural systems for which Gott's formula does not give accurate confidence intervals. In fact, it gives confidence intervals which are too narrow. Another way of stating this is that there are implicit assumptions made when one uses Gott's formula. Because one would need to have specific knowledge about the system to know whether or not these assumptions are satisfied, the method is not as powerful as previously thought.

Gott's formula

Image that time is separated into a doubly-infinite sequence of intervals:

$$\ldots, (\tau_{i-3}, \tau_{i-2}], (\tau_{i-2}, \tau_{i-1}], (\tau_{i-1}, \tau_i], (\tau_i, \tau_{i+1}], (\tau_{i+1}, \tau_{i+2}], \ldots \tag{1}$$

Assume that one finds themself T time units into a time period. Specifically, let the period have started at time τ_n, let the current time be t (thus $T = t - \tau_n$), let τ_{n+1} the ending time of the period (which one does not yet know), and let S be the remaining time of the time period, $\tau_{n+1} - t$. S is the quantity which one hopes to estimate.

The goal is to find a confidence interval for S, using only T as data. Gott's argument is as follows. Assume one wants to want to find a 50% confidence interval. There is no reason to think *a priori* that they occupy any specific point in the region (τ_n, τ_{n+1}). This is called the Copernican principle in time. A mathematical statement of this principle would be that one is assuming that the distribution of T is stationary. Based on this assumption, one can say that the probability that they lay in the middle one-half of the time period will be one-half. This is the region $\left[\tau_n + \frac{1}{4}(\tau_{n+1} - \tau_n), \tau_n + \frac{3}{4}(\tau_{n+1} - \tau_n)\right]$. This means that the amount of time remaining in the time period, S, will be with probability one-half between $1/3T$ and $3T$. This is because if one is at the point in time one-quarter into the time period, the future will be three times as long as the past, while if one is at the point three-quarters of the way into the time period, the future will be one-third as long as the past. With probability one-half, the length of the future will lie between these two values.

This is easily generalized to build a confidence interval of level $1 - \alpha$. If $T/(T+S)$ is $\alpha/2$, than $S = T(2-\alpha)/\alpha$. Likewise if $S/(T+S) = \alpha/2$ than $S = T\alpha/(2-\alpha)$. The $1 - \alpha$ confidence interval will be

$$\left(\frac{\alpha}{2-\alpha}T, \frac{2-\alpha}{\alpha}T\right). \tag{2}$$

Next I will build a mathematical model of the epochs and explore how Gott's formula works in a specific situation. I will show that in order for the formula to work, it is necessary that certain implicit assumptions are satisfied.

A mathematical model

Let time be partitioned as described above. Assume that the length of each interval is exponentially distributed. That is,
$P[\tau_n - \tau_{n-1} > T] = \exp[-\lambda_n T]$. Furthermore, assume each of the λ_n is a independent, identically distributed random variable which is exponentially distributed with mean one. That is, $P[\lambda_n > \theta] = \exp[-\theta]$.

In order to test the accuracy of the suggested confidence intervals, we need to compute $P[\tau_{n+1} - \tau_n > T + S \mid \tau_{n+1} - \tau_n > T]$. First, we note that

$$\begin{aligned} P[\tau_{n+1} - \tau_n > T] &= \int_0^\infty P[\tau_{n+1} - \tau_n > T \mid \lambda = \gamma] P(d\gamma) \\ &= \int_0^\infty e^{-\gamma T} e^{-\gamma}\, d\gamma = \frac{1}{T+1} \end{aligned} \tag{3}$$

where P is the density of the random variable λ. Likewise

$$P[\tau_{n+1} - \tau_n > S + T] = \frac{1}{S+T+1}, \tag{4}$$

and thus

$$\begin{aligned} P[\tau_{n+1} - \tau_n > T + S \mid \tau_{n+1} - \tau_n > T] &= \frac{P[\tau_{n+1} - \tau_n > S + T \,\cap\, \tau_{n+1} - \tau_n > T]}{P[\tau_{n+1} - \tau_n > T]} \\ &= \frac{P[\tau_{n+1} - \tau_n > S + T]}{P[\tau_{n+1} - \tau_n > T]} \\ &= \frac{T+1}{S+T+1}. \end{aligned} \tag{5}$$

Confidence intervals for S

Gott argues (as described above) that for a $1-\alpha$ confidence interval of S, the time remaining in the period, given that the time since the beginning of the period is T, one should use the interval $(S_l, S_h) = (T\alpha/(2-\alpha), T(2-\alpha)/\alpha)$. One can use (5) to check the accuracy of this claim. This is not a true $1-\alpha$ confidence interval as seen by computing

$$\begin{aligned} P[S_l < S < S_h] &= P[\tau_{n+1} - \tau_n > T + S_l \mid \tau_{n+1} - \tau_n > T] \\ &- P[\tau_{n+1} - \tau_n > T + S_h \mid \tau_{n+1} - \tau_n > T] \\ &= \frac{4T(T+1)(1-\alpha)}{(2T+2-\alpha)(2T+\alpha)} < 1 - \alpha. \end{aligned} \tag{6}$$

The problem here is not only that the interval is not a true confidence interval, but that it is always too *small*. The final inequality holds for all T, and for all α not equal to 0 or 1.

Another approach to building a confidence interval is to use information of the specific system. In this case, by using (5), one can build a $1-\alpha$ confidence interval by solving $(T+1)/(\hat{S}_h + T + 1) = \alpha/2$, and $1 - (T+1)/(\hat{S}_l + T + 1) = \alpha/2$. The confidence interval is $(\hat{S}_l, \hat{S}_h)$

$$= \left(\frac{\alpha}{2-\alpha}(T+1), \frac{2-\alpha}{\alpha}(T+1) \right). \tag{7}$$

Of course, the second confidence interval (7) is obtained through a entirely different approach. It is a true confidence interval for every T, while Gott's formula claims to build confidence intervals which work only for a random T, chosen from the stationary distribution of the system. In other words, if the stationary distribution of this process existed and had density $f(T)$, and furthermore if T were randomly chosen from this distribution, then the claim of Gott's formula would be that

$$\int_0^\infty \left(\frac{T+1}{\frac{\alpha}{2-\alpha}T + T + 1} - \frac{T+1}{\frac{2-\alpha}{\alpha}T + T + 1} \right) f(T)\, dt = 1 - \alpha. \tag{8}$$

However we know that for α not equal to 0 or 1,

$$\int_0^\infty \left(\frac{T+1}{\frac{\alpha}{2-\alpha}T + T + 1} - \frac{T+1}{\frac{2-\alpha}{\alpha}T + T + 1} \right) f(T)\, dt < (1-\alpha) \int_0^\infty f(T) dT < 1-\alpha. \tag{9}$$

It is not only for a fixed T, but also for T randomly chosen from a probability distribution that the confidence intervals constructed using Gott's formula are not accurate.

Further understanding of the problem at hand can be obtained by substituting $S = T$ into equation (5). By doing this one sees

$$P[\tau_{n+1} - \tau_n > 2\,T | \tau_{n+1} - \tau_n > T] = \frac{1+T}{1+2T}. \tag{10}$$

This says that for any T, the probability that one has to wait longer for the end of the epoch than they have already waited is *greater than one-half*! In other words, the probability that one is in the first half of the time period is greater then one-half, no matter what T they observe. This is a complete contradiction of the Copernican principle in time (and also of our intuition), and perhaps it leads us to understanding why in this case the method does not work. The problem arises because the expected future length of the time period, given the past length of the time period is infinite:

$$E[S \mid T] = \int_0^\infty S \frac{T+1}{(S+T+1)^2} dS = \infty \tag{11}$$

Furthermore, this implies that T does not have a stationary distribution, and the Copernican principle is explicitly violated.

In the traditional practice of statistics, confidence intervals are used to predict parameters, not random variables. Thus, one does not have to worry about finite expectation of the quantity one is estimating. However in this case we see that if the system is such that the first moment of the conditional length of the future based on the past is infinite, the method for constructing confidence intervals does not work.

Conclusion

Despite its elegance and power, there are difficulties in applying Gott's formula. The mathematical model which I used to describe the distribution of time periods is not so exotic as to be ruled out *a priori* as a description of a system about which nothing is known. It should be noted that it is not the case that such a system is so badly behaved that building confidence intervals is impossible. We did so in equation (7). Nor is the problem our example causes unique; there are many mathematical models describing time intervals in which the length of the intervals has an infinite first moment. Another example would be one in which the ends of the time period are given by the returns to zero of a symmetric random walk. Gott's formula will not work for any such system. Furthermore, Gott's formula is meant to be used as a completely non-parametric method. That is, one need not make any assumptions about the nature of the system they are studying. Here, however, we see that one is making the implicit assumption that the expectation of the future length of the time period conditioned on the past length of the time period is finite. Without knowing anything of the system under consideration, it is impossible to know whether this condition is satisfied.

References

[1] Gott , J.R. *Nature* **363** 315-319 (1993).

[2] Gott, J.R. in *Clusters, Lensing and the Future of the Universe* (Trimble, V. & Reisenegger, A., eds) 140 (Astron.Soc.Pacific Conf.Series, Vol 88, 1996).

[3] Landsberg, P.T., Dewynne, J.N. & Please, C.P. *Nature* **365**, 384 (1993).