Stochastic Aspects in Estimating the Probability of Producing Good Products by a System

— **Xubin Zeng** (xubin@halo.atmo.arizona.edu), Department of Atmospheric Sciences, The University of Arizona
Tucson, AZ 85721

Recently, Barkstrom (1995) applied a model, which is used in manufacturing systems engineering for machines that fail and have to be repaired, to estimate the probability of producing good data with an algorithm. In addition, he discussed the implications of this interesting model for EOS data production, and proposed four "brain teasers" for reader involvement at the end of the article.

Briefly, the first "brain teaser" is to either provide a more-detailed justification (than in Barkstrom) for Eq. 1 below or suggest an alternative form. The purpose of this short paper is to discuss the stochastic aspect of his first "brain teaser," which will also influence any answers to the second to fourth "brain teasers."

Following Barkstrom, the probability (q) of producing good data with an algorithm can be estimated by

$$q = \frac{1}{1 + T_r p} \qquad (1)$$

with

$$p = p_0 \, exp(-t / \lambda_p) \qquad (2)$$

where t is time, p_0 is the rate at which errors are discovered initially, λ_p is the error discovery lifetime, and $Tr = 1/r$ with r being the rate at which errors are fixed (corrected). Although the exponential decrease of p with time was considered in Eq. (2) in Barkstrom, T_r was taken as an empirical constant.

As mentioned by Barkstrom from his Earth Radiation Budget Experiment (ERBE) experience, different times were needed to repair different errors in ERBE algorithms. When working on numerical model development in the past few years, I have also had a similar experience: Initially, many errors occur but they are quite easy to fix; as time goes by, fewer errors are left but the mean time to fix them is usually longer. Usually, the model can be run without a floating-point error but gives unreasonable results. In that case, the fixing time can be short or long, largely depending upon the experience, talent, and luck of the researchers. Therefore, instead of assuming a constant T_r in Barkstrom, it may be more reasonable to assume

$$T_r = T_{r0} \, exp(t / \lambda_r) \zeta \qquad (3)$$

where λ_r is the "error repair time" (consistent with the definition of λ_p in Eq. (2)), and ζ is a random number with a mean value of unity. Using Eqs. (1)-(3), we obtain

$$q = \frac{1}{1 + T_{r0} p_0 \exp(-t / \lambda) \zeta} \qquad (4)$$

with

$$\frac{1}{\lambda} = \frac{1}{\lambda_p} - \frac{1}{\lambda_r} \qquad (5)$$

Therefore, if ζ is taken as unity, Eq. (4) is the same as Eq. (3) in Barkstrom except that λ is given in Eq. (5) instead of being λ_p in Barkstrom. In other words, without considering the stochastic effects, Eq. 3 in Barkstrom can also account for both the exponential decrease of p and the exponential increase of T_r with time.

When we consider the stochastic aspect of Eq. (4), we assume the probability density function of ζ

$$f(\zeta) = e^{-\zeta} \quad \text{for} \quad 0 < \zeta < \infty \qquad (6)$$

so that the expected, i.e., mean value $E(\zeta) = 1$.

Using Eqs. (4) and (6), we can obtain the expected (i.e., mean) value of q

$$E(q) = \int_{\circ}^{\infty} \frac{e^{-\zeta}}{1 + T_{r0} p_0 \exp(-t / \lambda) \zeta} d\zeta \qquad (7)$$

and the standard deviation of q, $S(q)$ comes from

$$S^2(q) = \int_{\circ}^{\infty} q^2(\zeta) e^{-\zeta} \; d\zeta - E^2(q) \qquad (8)$$

Equation (7) can be solved numerically in a computer; it can also be converted to a standard Exponential Integration (which is one of many special mathematical functions), and then solved using mathematical tables. We can also obtain from Eqs. (7)-(8)

$$S^2(q) = \frac{1}{p_0 T_{r0} \exp(-t/\lambda)} [1 - E(q)] - E^2(q) \qquad (9)$$

so that $S(q)$ can be easily computed.

Using the empirical values in Barkstrom, i.e., p_0 = 24 per year, T_{r0}= 0.5 year, and λ = 0.3 year, Figure 1 shows $E(q)$ and $S(q)$ as a function of time. The implication of Fig. 1 is that, depending on their experience and luck, and the complexity of the computer code, different EOS algorithm teams will spend different time periods to obtain good results. For instance, if we define t_{rep} to be the time period after which $E(q)$ is greater than 0.99 (as in Barkstrom) so that data reprocessing can start, then t_{rep}= 2.1 years for a team following the $E(q)$ curve (as in Barkstrom), t_{rep} = 1.4 years for a team following the $E(q) + S(q)$ curve, and t_{rep} = 2.3 years for a team following the $E(q)$-$S(q)$ curve. If data from various instruments on EOS satellites are needed for multidisciplinary studies, EOSDIS data users have to wait for a longer period of time than users using data from a single instrument only.

It is also seen from Fig. 1 that the standard deviation $S(q)$ is quite large during the first year with a peak at t = 0.5 year. This implies that the progress of different EOS algorithm teams could be quite different during the first year. Therefore, additional efforts should be made by various algorithm teams during the first year.

Note that $E(q)$ is slightly different from q in Barkstrom, and it can be proved mathematically that, with the probability density function in Eq. (6), $E(q)$ is always slightly larger than q with ζ = 1. However, this difference does not affect our discussions here. Note also that the functional forms assumed in Eqs. (3)-(4) are based on our experience; however, use of different

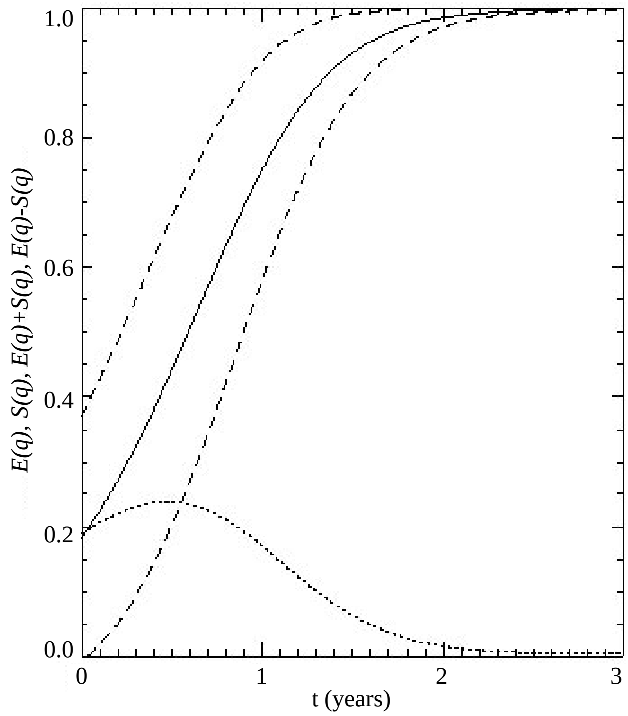

Figure 1. The mean value (E) and standard deviation (S) of q as a function of time, denoted by solid and dotted lines respectively. ($E(q) + S(q)$) and ($E(q) - S(q)$) are denoted by dashed lines above and below the solid line respectively.

functional forms with a stochastic component should not change our results qualitatively.

In summary, we have shown that Eq. (3) in Barkstrom can be used to account for both the exponential decrease of the error discovery rate with time and the exponential increase of the error repair rate with time. In addition, even with the same parameters in Eq. (4), depending on luck, different EOS algorithm teams will spend different time periods to obtain good data. This provides an answer to Barkstrom's first "brain teaser," and will also affect any answers to the three other "brain teasers."

Acknowledgement: This work was supported by NSF grant ATM-9419715.

References:

Barkstrom, B., 1995: The good, the bad, and the useful: Do things ever go right? ***The Earth Observer***, **7** , 46-49.

■