Ancient TL

www.ancienttl.org · ISSN: 2693-0935

Galbraith, R., 2002. *A note on the variance of a background-corrected OSL count.* Ancient TL 20(2): 49-51. https://doi.org/10.26034/la.atl.2002.348

A note on the variance of a background-corrected OSL count

Rex Galbraith
Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, U.K.

(Received 14 May 2002; in final form 25 October 2002)

It is common practice to calculate the relative standard error of a background-corrected optically stimulated luminescence (OSL) count by assuming Poisson errors. This note corrects a formula given by Banerjee *et al.* (2000) and suggests alternative formulae for use when the variation in background counts is larger than that implied by the Poisson distribution. For moderately bright samples, the contribution to the relative standard error from estimating the background rate is small, whichever formula is used.

The usual scenario is as follows. Optical stimulation of an aliquot of quartz produces a series of counts - a number of recorded photons for each of N equal length consecutive time intervals (channels). For example, Banerjee *et al.* (2000) used a stimulation period of 60 s with counts in N = 250 channels each lasting 0.24 s. The OSL "signal" is measured from the total count in the first n channels minus an estimate of the contribution to this count from background sources. Often n is taken to be quite small, for example n = 5, corresponding to the first 1.2 s of stimulation. The background emission rate is assumed to be constant over the whole 60 s, and is estimated from counts near the end of this period, where the contribution from the signal is assumed to be negligible.

Mathematically, the above may be expressed as follows. Let y_i denote the OSL count from channel i, for i = 1, 2, …, N, and let $Y_0 = \sum_{i=1}^{n} y_i$ be the total count over the first n channels. Write

$$Y_0 = S_0 + B_0$$

where S_0 and B_0 are the contributions to Y_0 from the signal (or source of interest) and background respectively. Of course S_0 and B_0 are not observed directly. Assume that S_0 and B_0 are independent random quantities with expectations μ_S and μ_B, and variances σ^2_S and σ^2_B, respectively. Then the observed count Y_0 will have expectation $\mu_S + \mu_B$ and variance $\sigma^2_S + \sigma^2_B$. An estimate of the signal μ_S is thus obtained by subtracting an estimate of μ_B from Y_0, i.e.,

$$\hat{\mu}_S = Y_0 - \hat{\mu}_B$$

We want to calculate the relative standard error of this estimate.

An estimate of μ_B is usually obtained from the average OSL count over the last m channels, for some suitable m chosen so that the contribution from the signal is negligible. It is useful to choose m be a multiple of n: let m = nk, say. For example, Banerjee *et al.* (2000) used the last m = 25 channels (6 s) of the series, corresponding to k=5 when n=5. Then let Y_1, Y_2, …, Y_k denote the total counts in the last k sets of n channels, i.e.,

$$Y_j = \sum_{i=N-jn+1}^{N-jn+n} y_i$$

for j = 1, 2, …, k. Thus Y_1, Y_2, …, Y_k are all counts over n channels (the same as for Y_0) and we assume that they are independent random quantities from the same distribution as that of B_0 (i.e., the signal is negligible). In particular, each has expectation μ_B and variance σ^2_B. The estimate of μ_B may then be written as

$$\hat{\mu}_B = \overline{Y} = \frac{1}{k}\sum_{j=1}^{k} Y_j$$

and this has variance $\mathrm{var}(\hat{\mu}_B) = \sigma_B^2 / k$ Hence the variance of the estimated signal (corrected for background) is

$$\mathrm{var}(\hat{\mu}_S) = \mathrm{var}(Y_0) + \mathrm{var}(\hat{\mu}_B) = \sigma_S^2 + \sigma_B^2 + \sigma_B^2 / k \qquad (1)$$

and the relative standard error is

$$rse(\hat{\mu}_s) = \frac{\sqrt{\sigma_S^2 + \sigma_B^2 + \sigma_B^2 / k}}{\mu_S} \quad (2)$$

In order to calculate this relative standard error in practice, we need estimates of σ_S^2 and σ_B^2 in addition to the estimate of μ_S.

In the usual case where $Y_0, Y_1 \ldots, Y_k$ are assumed to have Poisson distributions, $\sigma_S^2 = \mu_S$ and $\sigma_B^2 = \mu_B$. Then (1) becomes

$$\text{var}(\hat{\mu}_S) = \mu_S + \mu_B + \mu_B / k$$

which may be estimated as $Y_0 - \bar{Y} + \bar{Y} + \bar{Y}/k = Y_0 + \bar{Y}/k$. Substituting these estimates into (2) gives the following estimated relative standard error:

$$rse(\hat{\mu}_s) = \frac{\sqrt{Y_0 + \bar{Y}/k}}{Y_0 - \bar{Y}} \quad (3)$$

This differs slightly from the formula on page 833 of Banerjee *et al.* (2000), where the second term in the numerator is equivalent to $2\bar{Y}/k$. The above argument shows that the factor 2 should not be there.

A drawback with equation (3) in practice is that there is sometimes evidence that the background counts do not have a Poisson distribution, but are over-dispersed (e.g., Galbraith *et al.*, 1999, p 348). For example, the variance

$$s_Y^2 = [\,1/(k-1)]\,\Sigma_{j=1}^{k}\,(Y_j - \bar{Y})^2$$

may be substantially larger than the mean count $\bar{Y}$ (see below). Then we may write

$$\sigma_B^2 = \mu_B + \sigma^2$$

for some positive value of σ^2 to be estimated. An obvious estimate is

$$\hat{\sigma}^2 = s_Y^2 - \bar{Y} \quad (4)$$

provided this is positive. But there is a drawback with this too: in order to be confident that the contribution to $\bar{Y}$ from the signal is negligible, it might be necessary to use a quite small value of k (e.g., k=5 as above), so that s_Y^2 will be based on a small number of degrees of freedom. A more reliable estimate may be obtained by pooling the background variances for several series. For example, for four series of stimulation with background means $\bar{Y}_1$, $\bar{Y}_2$, $\bar{Y}_3$, $\bar{Y}_4$ and variances s_{Y1}^2, s_{Y2}^2, s_{Y3}^2, s_{Y4}^2, each with k–1 degrees of freedom, one may use

$$\hat{\sigma}^2 = \frac{1}{4}(s_{Y1}^2 + s_{Y2}^2 + s_{Y3}^2 + s_{Y4}^2) - \frac{1}{4}(\bar{Y}_1 + \bar{Y}_2 + \bar{Y}_3 + \bar{Y}_4) \quad (5)$$

i.e., the average variance minus the average background count for the four series. This pooled estimate of over-dispersion could be used for each series, while at the same time using separate estimates of background level.

It is not so straightforward to obtain a corresponding estimate of σ_S^2 because the expected counts change rapidly at the start of the stimulation period. But there is perhaps a case for assuming that S_0 does have a Poisson distribution, while B_0 does not. The former comes from pure OSL emissions while the latter comes from other sources such as scattered light and instrument noise, which may not exhibit Poisson variation. Then we still have $\sigma_S^2 = \mu_S$ and the resulting estimated relative standard error is

$$rse(\hat{\mu}_S) \approx \frac{\sqrt{Y_0 + \bar{Y}/k + \hat{\sigma}^2(1 + 1/k)}}{Y_0 - \bar{Y}} \quad (6)$$

This formula will agree closely with (3) when $\hat{\sigma}^2$ is small, but may be preferable when $\hat{\sigma}^2$ is large.

To get a feel for the numerical consequences, suppose $Y_0 = 12500$ and $\bar{Y} = 50$, with n=5 and k=5. These numbers are comparable with the "bright" sample in Banerjee *et al.* (2000). Suppose also that $\hat{\sigma}^2 = 75$, corresponding to a reasonably substantial amount of over-dispersion ($\sigma_B^2/\mu_B \approx 2.5$). Then the relative standard errors from (3) and (6) are 0.00898 and 0.00902, respectively, which are practically equal. The corresponding absolute standard errors are 111.8 and 112.2. Indeed, if we treated the background as being known *exactly*, the relative standard error of $\hat{\sigma}^2$ would be 0.00896, also practically the same. But for a weak signal, with $Y_0 = 200$ and the same $\bar{Y}$ and $\hat{\sigma}^2$ as above, the two relative standard errors are 0.0966 and 0.1155, and the two absolute standard errors are 14.5 and 17.2. In general, when the signal is weak, equation (6) may give a somewhat larger relative standard error than (3). But when the signal is strong, both equations give similar answers and in fact the error in estimating the background is practically negligible.

An alternative assumption when there is over-dispersion is that neither S_0 nor B_0 have Poisson distributions, but that the ratio of the variance to the mean is the same for each, i.e., $\sigma_S^2/\mu_S = \sigma_B^2/\mu_B$. It is perhaps hard to think of a physical justification for this: it would imply a multiplicative error mechanism that affected all counts from whatever source. So it would presumably be error associated with the measurement process rather than the process producing the counts. In any case, this would lead to the following estimated relative standard error:

$$rse(\hat{\mu}_S) \approx \sqrt{1+\frac{\hat{\sigma}^2}{\bar{Y}}} \times \frac{\sqrt{Y_0 + \bar{Y}/k}}{Y_0 - \bar{Y}} \qquad (7)$$

where $\hat{\sigma}^2$ is given by (4) or (5). Here equation (3) is multiplied by a factor corresponding to $\sqrt{\sigma_B^2/\mu_B}$. In the above examples this factor would $\sqrt{2.5} \approx 1.6$. In general, this formula is more conservative than (3) or (6).

Some further remarks may be useful. Firstly, there is a simple statistical test for assessing whether the counts Y_1, Y_2, …, Y_k vary consistently with a Poisson distribution. One calculates the Poisson index of dispersion

$$I = (k-1)s_Y^2 / \bar{Y}$$

and assesses its significance from the χ^2 distribution with k−1 degrees of freedom (see for example Kotz and Johnson, 1987, p 25). The quantity I is akin to a χ^2 statistic, a significantly large value of I being evidence of over-dispersion, i.e., the ratio $s_Y^2 / \bar{Y}$ is too large to be consistent with $\sigma_B^2/\mu_B = 1$. Of course there are various ways to test whether data agree with a Poisson distribution; this is a useful method when there is only a small number of counts.

Secondly, estimating a palaeodose typically uses products and ratios of background-corrected OSL counts. Then the approximate relative standard error of the product or ratio is simply obtained by combining the individual relative standard errors in quadrature. For example, for the simplest form of the single-aliquot regenerative-dose protocol,

$$\text{palaeodose} = \frac{s_n}{s_r} \times \frac{t_r}{t_n} \times \text{regenerative dose},$$

where s_n, t_n, s_r and t_r are estimated from background-corrected counts arising from optical stimulation of the natural dose, a subsequent test dose, a regenerative dose and its corresponding test dose, respectively. Then the relative standard error of the palaeodose estimate (assuming the regenerative dose is known exactly) is just the square root of the sum of the squared relative standard errors of the estimates of s_n, t_n, s_r and t_r.

Thirdly, subtracting an estimated background level from a weak signal can produce inappropriate estimates. An alternative approach is to estimate the signal *in the presence of* the background using statistical models (e.g., Galbraith *et al.,* 1999).

Finally, when estimating palaeodoses or other OSL parameters from several aliquots of quartz, whether using single- or multiple-aliquot methods, there are usually other sources of variation to account for in addition to those reflected in (3), (6) or (7). These may be more substantial and work on estimating them is in progress.

I thank John Prescott, Bert Roberts and Andrew Murray for useful comments on an earlier version of this note.

References

Banerjee, D., Bøtter-Jensen, L. & Murray, A.S. (2000) Retrospective dosimetry: estimation of the dose to quartz using the single-aliquot regenerative-dose protocol. *Applied Radiation and Isotopes,* **52,** 831-844.

Galbraith, R.F., Roberts, R.G., Laslett, G.M., Yoshida, H. and Olley, J.M. (1999) Optical dating of single and multiple grains of quartz from Jinmium rock shelter, Northern Australia: Part I, Experimental design and statistical models. *Archaeometry,* **41,** 338-364.

Kotz,S. & Johnson, N.L., Editors (1987) Poisson Index of Dispersion. *Encyclopedia of Statistical Sciences,* volume 7, Wiley 1987.

Reviewer

John Prescott