Greenwood's formula puts a standard error on the Kaplan-Meier estimator using the delta-method. At any particular time t with a failure, let N_t be the number of subjects on test "at time $t-$," that is, just before time t. The probability of surviving from $t-$ to $t+$ is estimated as X_t/N_t, where X_t is the number who survive from $t-$ to $t+$. The Kaplan-Meir estimator is

$$T \to \prod_{t<T} \frac{X_t}{N_t}. \tag{1}$$

The X_t are modeled as independent binomial $B(N_t, p_t)$ variables. Independence is clearly wrong, randomness of failure times is ignored, and hidden randomness—absence of failure between observed failure times—is ignored. Finite-sample, this is no good. Asymptotically, under conditions, might be fine.

Anyway, we make the modeling assumptions. Let $\hat{K}$ be the product, with expected value K. So

$$\frac{\hat{K}}{K} = \prod \frac{X_t}{N_t p_t} = \prod \left(1 + \frac{X_t - N_t p_t}{N_t p_t}\right) \approx 1 + \sum \frac{X_t - N_t p_t}{N_t p_t} \tag{2}$$

provided the $N_t p_t$ are all large. So

$$\operatorname{var}\left(\frac{\hat{K}}{K}\right) \approx \sum \frac{1 - p_t}{N_t p_t} \approx \sum \frac{1 - \hat{p}_t}{N_t \hat{p}_t} \tag{3}$$

where $\hat{p}_t = X_t/N_t$. Notice that p_t is the survival probability and $\hat{p}_t$ is the estimated survival probability. Thus,

$$\operatorname{var}\big(\hat{K}\big) \approx K^2 \sum \frac{1 - \hat{p}_t}{N_t \hat{p}_t} \approx \hat{K}^2 \sum \frac{1 - \hat{p}_t}{N_t \hat{p}_t} \tag{4}$$

This is Greenwood's formula.

Under suitable conditions, asymptotically, $\hat{K}/K$—so also $K/\hat{K}$—is nearly distributed as

$$N\left(1, \sum \frac{1 - \hat{p}_t}{N_t \hat{p}_t}\right) \tag{5}$$

which gives confidence intervals.

It may speed up convergence to work on a logarithmic scale, and this gives another standard approximation:

$$\log \hat{K} = \log K + \sum \log \left(1 + \frac{X_t - N_t p_t}{N_t p_t}\right) \approx \log K + \sum \frac{X_t - N_t p_t}{N_t p_t} \tag{6}$$

so

$$\operatorname{var}\big(\log \hat{K}\big) \approx \sum \frac{1 - \hat{p}_t}{N_t \hat{p}_t} \tag{7}$$

Under suitable conditions, asymptotically,

$$\log \hat{K} = \log K + \zeta_n, \text{ where } \zeta_n \sim N\left(0, \sum \frac{1 - \hat{p}_t}{N_t \hat{p}_t}\right) \tag{8}$$

This is another way to get confidence intervals.

Sometimes, two logs are used:

$$\log(-\log \hat{K}) \tag{9}$$

More specifically, we start from (8), substitute $\log \hat{K} = \log K + \zeta_n = \log K(1 + \zeta_n/\log K)$ into (9), and use the delta-method. This may further speed up convergence (Borgan and Liestrøl, 1990).

Reference

Major Greenwood, Jr. (1926). The Natural Duration of Cancer. *Reports of Public Health and Related Subjects*, Vol. 33, HMSO, London.

Ørnulf Borgan and K Liestrøl (1990). A note on confidence intervals and bands for the survival curve based on transformations. Scand J Statist 17: 35–41.

Per Kragh Andersen and Niels Keiding, editors (2006). *Survival and Event History Analysis*. Wiley.