The Delta Method

Often one has one or more MLEs ($\hat{\theta}_i$) and their estimated, conditional sampling variance-covariance matrix. However, there is interest in some function of these estimates. The question is, “what is the variance of this new quality?" Clearly, the θ_i are random variables, thus the new quantity is a random variable and has some sampling variance as a measure of its precision. Consider the mean life span for adult animals after they have been tagged. An estimator of mean life span, $\bar{\ell}$, is

$$\bar{\ell} = \frac{-1}{\log_e(\hat{S})} \ .$$

Assume that the MLE $\hat{S}$ and its conditional sampling variance are available. So, what is $\text{var}(\bar{\ell})$? This question can be addressed by what is loosely called the delta method.

Transformations of One Variable

For the case where there is a simple, nonlinear transformation (as above), the procedure is simple (if one recalls the calculus or knows how to use programs such as *DERIVE* or *MAPLE*). That is,

$$\widehat{\text{var}}(\bar{\ell}) = \left(\frac{\partial \bar{\ell}}{\partial \hat{S}}\right)^2 \cdot \widehat{\text{var}}(\hat{S}).$$

The sampling variance of $\bar{\ell}$ is just the squared partial derivative of $\bar{\ell}$ with respect to $\hat{S}$ times the sampling variance of $\hat{S}$.

Lets check this out in a case where we know the answer. Assume we have an estimate of density $\hat{D}$ and its conditional sampling variance, $\widehat{\text{var}}(\hat{D})$. We want to multiply this by some constant c to make it comparable with other values from the literature. Thus, we want $\hat{D}_s$ = $c\hat{D}$ and $\text{var}(\hat{D}_s)$. From simple statistics, we know that

$$\text{var}(\hat{D}_s) = c^2 \cdot \text{var}(\hat{D}).$$

The delta method gives

$$\text{var}(\hat{D}_s) = \left(\frac{\partial \hat{D}_s}{\partial \hat{D}}\right)^2 \cdot \widehat{\text{var}}(\hat{D})$$

$$= \text{c}^2 \cdot \text{var}(\hat{D}).$$

Another example is a known number of fish N and an average weight ($\hat{\mu}_w$) and its variance. If you want biomass, then $\hat{B} = N \cdot \hat{\mu}_w$ and the variance of $\hat{B}$ is $N^2 \cdot \widehat{\text{var}}(\hat{\mu}_w)$.

Some other results:

$$\widehat{\text{var}}(\hat{\theta}/c) = (1/c)^2 \cdot \widehat{\text{var}}(\hat{\theta}),$$

where *c* is a constant.

The variance of a sum is the sum of the variances of the elements are independent. Thus,

$$\widehat{\text{var}}(\sum \hat{\theta}_i) = \sum \widehat{\text{var}}(\hat{\theta}_i).$$

From this, can you determine the variance of a mean?

If the terms are dependent, then

$$\widehat{\text{var}}(\sum \hat{\theta}_i) = \sum \widehat{\text{var}}(\hat{\theta}_i) + \sum_{i \neq} \sum_{j} \widehat{\text{cov}}(\hat{\theta}_i\ \hat{\theta}_j).$$

If this looks messy, it is merely the sum of all the elements in the sampling variance-covariance matrix.

Note, the variance of a difference is just,

$$\widehat{\text{var}}(\hat{\theta}_i - \hat{\theta}_j) = \widehat{\text{var}}(\hat{\theta}_i) + \widehat{\text{var}}(\hat{\theta}_j) - 2\widehat{\text{cov}}(\hat{\theta}_i\ \hat{\theta}_j).$$

What if the estimates were independent?

The delta method works well, particularly if the coefficients of variation are “small." It is a very handy tool, but not computer intensive, like the bootstrap.

Another example, one has an MLE $\hat{\gamma}$ and $\widehat{\text{var}}(\hat{\gamma})$, but makes the transformation,

$$\hat{\psi} = e^{\hat{\gamma}^2} = \exp(\hat{\gamma}^2).$$

Then,

$$\widehat{\text{var}}(\hat{\psi}) = \left(\frac{\partial \hat{\psi}}{\partial \hat{\gamma}}\right)^2 \cdot \widehat{\text{var}}(\hat{\gamma}).$$

Or, one could write this as

$$\widehat{\text{var}}(\hat{\psi}) = \left(\frac{\partial \hat{\psi}}{\partial \hat{\gamma}}\right) \cdot \widehat{\text{var}}(\hat{\gamma}) \cdot \left(\frac{\partial \hat{\psi}}{\partial \hat{\gamma}}\right),$$

and this will be useful for extensions (below).
It turns out that the partial of ψ wrt γ is $2\gamma \cdot \exp(\gamma^2)$, thus

$$\hat{\text{v}}\text{ar}(\hat{\psi}) = \left(2\gamma \cdot \exp(\gamma^2)\right) \cdot \hat{\text{v}}\text{ar}(\hat{\gamma}) \cdot \left(2\gamma \cdot \exp(\gamma^2)\right) .$$

This appears slightly nasty, but is certainly computable by hand, if need be. Check these results with program *DERIVE*. Try some transformation between the instantaneous rate $\hat{r} = -\log_e(\hat{S})$, where the MLE $\hat{S}$ is known, as is its variance. What is the variance of the estimated instantaneous rate?

Here is a more challenging example (still with only one parameter estimate $\hat{\theta}$). Based on a fairly large data set, you compute the MLE of the parameter θ and its conditional sampling variance: $\hat{\theta}$ and $\hat{\text{v}}\text{ar}(\hat{\theta})$. In ecotoxicology one might want a new quantity λ defined as

$$\hat{\lambda} = \log_e \left(\frac{\hat{\theta}^4}{(3\hat{\theta} - 4)^2} \right) .$$

This estimator of λ involves only one random variable ($\hat{\theta}$), thus the variance is approximately,

$$\hat{\text{v}}\text{ar}(\hat{\lambda}) = \left(\frac{\partial\hat{\lambda}}{\partial\hat{\theta}_i} \right) \cdot \hat{\text{v}}\text{ar}(\hat{\theta}) \cdot \left(\frac{\partial\hat{\lambda}}{\partial\hat{\theta}_i} \right) .$$

Proceed by writing $\hat{\lambda}$ as

$$\hat{\lambda} = \log_e(\hat{\theta}^4) - \log_e(3\hat{\theta} - 4)^2$$

$$= 4\log_e(\hat{\theta}) - 2\log_e(3\hat{\theta} - 4),$$

then the derivative wrt $\hat{\theta}$ is

$$\frac{4}{\hat{\theta}} - \frac{2 \cdot 3}{3\hat{\theta} - 4} .$$

Finally,

$$\hat{\text{v}}\text{ar}(\hat{\lambda}) = \left(\frac{4}{\hat{\theta}} - \frac{2 \cdot 3}{3\hat{\theta}-4}\right) \hat{\text{v}}\text{ar}(\hat{\theta}) \left(\frac{4}{\hat{\theta}} - \frac{2 \cdot 3}{3\hat{\theta}-4}\right) .$$

Transformations of Several Variables

This is the interesting case where the delta method is very useful in estimating approximate sampling variances of functions of random variables. Now, assume you compute $\mathcal{Y}$ as some linear or nonlinear function of $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$, and $\hat{\theta}_4$. You have the 4 MLEs and their 4x4 estimated variance-covariance matrix $\hat{\sum}$. The general form for the variance of $\mathcal{Y}$ is

$$\hat{\text{v}}\text{ar}(\hat{\mathcal{Y}}) = \left(\frac{\partial\hat{\mathcal{Y}}}{\partial\hat{\theta}_i}\right) \cdot \hat{\sum} \cdot \left(\frac{\partial\hat{\mathcal{Y}}}{\partial\hat{\theta}_i}\right)^T ,$$

where the first term on the RHS is the row vector with the partials of $\mathcal{Y}$ wrt $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$, and $\hat{\theta}_4$, respectively and the final term on the RHS is a column vector of the partials of $\mathcal{Y}$ wrt $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$, and $\hat{\theta}_4$, respectively. The row vector contains

$$\left(\frac{\partial\hat{\mathcal{Y}}}{\partial\hat{\theta}_1}\right), \left(\frac{\partial\hat{\mathcal{Y}}}{\partial\hat{\theta}_2}\right), \left(\frac{\partial\hat{\mathcal{Y}}}{\partial\hat{\theta}_3}\right), \left(\frac{\partial\hat{\mathcal{Y}}}{\partial\hat{\theta}_4}\right),$$

while the column vector contains

$$\left(\frac{\partial\hat{\mathcal{Y}}}{\partial\hat{\theta}_1}\right)$$

$$\left(\frac{\partial\hat{\mathcal{Y}}}{\partial\hat{\theta}_2}\right)$$

$$\left(\frac{\partial\hat{\mathcal{Y}}}{\partial\hat{\theta}_3}\right)$$

$$\left(\frac{\partial\hat{\mathcal{Y}}}{\partial\hat{\theta}_4}\right) .$$

Of course, the variance-covariance matrix is 4x4 so the matrix equation can be computed. The general result to remember is

$$\hat{\text{v}}\text{ar}(\hat{\mathcal{Y}}) = \left(\frac{\partial\hat{\mathcal{Y}}}{\partial\hat{\theta}_i}\right) \cdot \sum \cdot \left(\frac{\partial\hat{\mathcal{Y}}}{\partial\hat{\theta}_i}\right)^T .$$

Estimation of “Reporting Rate"

In some cases animals are tagged or banded to estimate a “reporting rate" – the proportion of banded animals reported, given that they were killed and retrieved by a hunter or angler. Thus, N_c animals are tagged with normal (Control) tags and, of these, R_c are recovered the first year following release. The recovery rate of control animals is merely R_c/N_c and we denote this as f_c.

Another group of animals, of size N_r, are tagged with REWARD tags; these tags indicate that $50 will be given to people reporting these special tags. It is assumed that all such tags will be reported, thus these serve as a basis for comparison and the estimation of a reporting rate. The recovery rate for the reward tagged animals is merely R_r/N_r, where R_r is the number of recoveries of reward-tagged animals the first year following release. We denote this recovery rate as f_r.

The estimator of the reporting rate is a ratio of the recovery rates and we denote this a ψ. Thus, $\hat{\psi} = \hat{f}_c/\hat{f}_r$. Both recovery rates are binomials, thus

$$\hat{\text{v}}\text{ar}(\hat{f}_c) = \hat{f}_c(1-\hat{f}_c)/N_c \quad \text{and } \hat{\text{v}}\text{ar}(\hat{f}_r) = \hat{f}_r(1-\hat{f}_r)/N_r .$$

The samples are independent, thus $\text{cov}(f_c, f_r) \equiv 0$ and the sampling variance-covariance is diagonal.

First, we need the derivatives of ψ wrt f_c and f_r:

$$\frac{\partial\hat{\psi}}{\partial\hat{f}_c} = \frac{1}{f_r} \qquad\qquad \frac{\partial\hat{\psi}}{\partial\hat{f}_r} = -\frac{f_c}{f_r^2} ,$$

then,

$$\hat{\text{v}}\text{ar}(\hat{\psi}) = \left[\frac{1}{\hat{f}_r}, \quad -\frac{\hat{f}_c}{\hat{f}_r^2}\right] \begin{bmatrix} \hat{\text{v}}\text{ar}(\hat{f}_c) & 0 \\ 0 & \hat{\text{v}}\text{ar}(\hat{f}_r) \end{bmatrix} \begin{bmatrix} \frac{1}{f_r} \\ -\frac{f_c}{f_r^2} \end{bmatrix}$$

This matrix equation can be simplified,

$$\hat{\text{v}}\text{ar}(\hat{\psi}) = \frac{1}{(\hat{f}_r)^2} \cdot \hat{\text{v}}\text{ar}(\hat{f}_c) + \frac{(\hat{f}_c)^2}{(\hat{f}_r)^4} \cdot \hat{\text{v}}\text{ar}(\hat{f}_r) .$$

If you want a numerical example, try N_c = 1,000, R_c = 80 and N_r = 500, R_r = 60. What is $\hat{\psi}$ and its sampling variance. When the covariances are 0, the use of the delta method is easier.

A More Complicated Example

Seber (1973:254) developed a tagging model to allow MLEs of survival S and a sampling rate (λ), both assumed constant over years and ages and sexes. To be consistent with models developed by Brownie et al. (1985), we wish to transform λ into another version of the sampling probability, $f = (1–S)\lambda$. [The notation in this example and the previous example are unrelated.]

You have $\hat{S}$, $\hat{\text{v}}\text{ar}(\hat{S})$, $\hat{\lambda}$, and $\hat{\text{v}}\text{ar}(\hat{\lambda})$. To find $\hat{\text{v}}\text{ar}(\hat{f})$, we use the delta method. The variance-covariance matrix is

$$\hat{\Sigma} = \begin{bmatrix} \hat{\text{v}}\text{ar}(\hat{S}) & \hat{\text{c}}\text{ov}(\hat{S}, \hat{\lambda}) \\ \hat{\text{c}}\text{ov}(\hat{\lambda}, \hat{S}) & \hat{\text{v}}\text{ar}(\hat{\lambda}) \end{bmatrix} .$$

Transformations: $S \rightarrow S$ and $(1–S)\lambda \rightarrow f$. The partial derivatives are,

$$\frac{\partial \hat{S}}{\partial \hat{S}} = 1 \qquad \frac{\partial \hat{S}}{\partial \hat{\lambda}} = 0,$$

$$\frac{\partial f}{\partial S} = -\lambda = -f/(1-S) \qquad \frac{\partial f}{\partial \lambda} = 1-S.$$

Finally, the sampling variance-covariance matrix of S and f is

$$\begin{bmatrix} 1 & 0 \\ -f/(1-S) & (1-S) \end{bmatrix} \hat{\Sigma} \begin{bmatrix} 1 & -f/(1-S) \\ 0 & (1-S) \end{bmatrix} .$$

Thus,

$$\widehat{\mathrm{cov}}(\hat{S}, \hat{f}) = -\hat{f}/(1-\hat{S}) \cdot \widehat{\mathrm{var}}(\hat{S}) + (1-S) \cdot \widehat{\mathrm{cov}}(\hat{S}, \hat{\lambda})$$

and

$$\widehat{\mathrm{var}}(\hat{f}) = (\hat{f}/(1-\hat{S}))^2 \cdot \widehat{\mathrm{var}}(\hat{S}) - 2\hat{f} \cdot \widehat{\mathrm{cov}}(\hat{S}, \hat{\lambda}) + (1-\hat{S})^2 \cdot \widehat{\mathrm{var}}(\hat{\lambda}).$$