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 \( \theta_i \) 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, \( \hat{\ell} \), is

\[
\hat{\ell} = \frac{-1}{\log e(S)}. 
\]

Assume that the MLE \( \hat{S} \) and its conditional sampling variance are available. So, what is \( \text{var}(\hat{\ell}) \)? 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 \textit{DERIVE} or \textit{MAPLE}). That is,

\[
\hat{\text{var}}(\hat{\ell}) = \left( \frac{\partial \hat{\ell}}{\partial \hat{S}} \right)^2 \cdot \hat{\text{var}}(\hat{S}).
\]

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

Let’s 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, \( \hat{\text{var}}(\hat{D}) \). 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

\[
\text{var}(\hat{D}_s) = \left( \frac{\partial \hat{D}_s}{\partial \hat{D}} \right)^2 \cdot \hat{\text{var}}(\hat{D}) = c^2 \cdot \text{var}(\hat{D}).
\]

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 \( N^2 \cdot \hat{\text{var}}(\hat{\mu}_w) \).
Some other results:

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

where \( c \) is a constant.

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

\[ \hat{\text{var}}(\sum \hat{\theta}_i) = \sum \hat{\text{var}}(\hat{\theta}_i). \]

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

If the terms are dependent, then

\[ \hat{\text{var}}(\sum \hat{\theta}_i) = \sum \hat{\text{var}}(\hat{\theta}_i) + \sum \sum \hat{\text{cov}}(\hat{\theta}_i \hat{\theta}_j). \]

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,

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

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 \( \hat{\text{var}}(\hat{\gamma}) \), but makes the transformation,

\[ \hat{\psi} = e^{\hat{\delta}^2} = \exp(\hat{\gamma}^2). \]

Then,

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

Or, one could write this as

\[ \hat{\text{var}}(\hat{\psi}) = \left( \frac{\partial \hat{\psi}}{\partial \gamma} \right) \cdot \hat{\text{var}}(\hat{\gamma}) \cdot \left( \frac{\partial \hat{\psi}}{\partial \gamma} \right), \]
and this will be useful for extensions (below).

It turns out that the partial of $\psi$ wrt $\gamma$ is $2\gamma \cdot \exp(\gamma^2)$, thus

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

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 $\theta$ and its conditional sampling variance: $\hat{\theta}$ and $\text{var}(\hat{\theta})$. In ecotoxicology one might want a new quantity $\lambda$ defined as

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

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

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

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

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

Finally,
\[
\hat{\text{var}}(\lambda) = \left( \frac{4}{\hat{\theta}} - \frac{2.3}{3\hat{\theta}^2 - 4} \right) \hat{\text{var}}(\hat{\theta}) \left( \frac{4}{\hat{\theta}} - \frac{2.3}{3\hat{\theta}^2 - 4} \right).
\]

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 \( Y \) 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 \( \hat{\Sigma} \). The general form for the variance of \( Y \) is

\[
\hat{\text{var}}(Y) = \left( \frac{\partial Y}{\partial \hat{\theta}_i} \right) \cdot \hat{\Sigma} \cdot \left( \frac{\partial Y}{\partial \hat{\theta}_i} \right)^T,
\]

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

\[
\begin{pmatrix}
    \frac{\partial Y}{\partial \hat{\theta}_1} \\
    \frac{\partial Y}{\partial \hat{\theta}_2} \\
    \frac{\partial Y}{\partial \hat{\theta}_3} \\
    \frac{\partial Y}{\partial \hat{\theta}_4}
\end{pmatrix},
\]

while the column vector contains

\[
\begin{pmatrix}
    \frac{\partial Y}{\partial \hat{\theta}_1} \\
    \frac{\partial Y}{\partial \hat{\theta}_2} \\
    \frac{\partial Y}{\partial \hat{\theta}_3} \\
    \frac{\partial Y}{\partial \hat{\theta}_4}
\end{pmatrix}.
\]
Of course, the variance-covariance matrix is 4x4 so the matrix equation can be computed. The general result to remember is

\[
\hat{\text{var}}(\hat{\mathbf{y}}) = \left( \frac{\partial \hat{\mathbf{y}}}{\partial \hat{\theta}_i} \right) \cdot \sum \cdot \left( \frac{\partial \hat{\mathbf{y}}}{\partial \hat{\theta}_i} \right)^T .
\]

**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 \( R/N_c \) 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 \( R/N_r \), where \( R \) 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 as \( \psi \). Thus, \( \hat{\psi} = \hat{f}_c/\hat{f}_r \). Both recovery rates are binomials, thus

\[
\hat{\text{var}}(\hat{f}_c) = \hat{f}_c (1-\hat{f}_c)/N_c \quad \text{and} \quad \hat{\text{var}}(\hat{f}_r) = \hat{f}_r (1-\hat{f}_r)/N_r .
\]

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 \( \psi \) wrt \( f_c \) and \( f_r \):

\[
\frac{\partial \psi}{\partial f_c} = \frac{1}{f_r} \quad \frac{\partial \psi}{\partial f_r} = -\frac{f_c}{f_r^2} ,
\]

then,
This matrix equation can be simplified,

\[ \hat{\text{var}}(\psi) = \frac{1}{(\hat{f}_r)^2} \cdot \hat{\text{var}}(\hat{f}_e) + \frac{(\hat{f}_r)^2}{(\hat{f}_r)^4} \cdot \hat{\text{var}}(\hat{f}_r). \]

If you want a numerical example, try \( N_c = 1,000, R_c = 80 \) and \( N_r = 500, R_r = 60 \). What is \( \psi \) 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 \( (\lambda) \), both assumed constant over years and ages and sexes. To be consistent with models developed by Brownie et al. (1985), we wish to transform \( \lambda \) into another version of the sampling probability, \( f = (1-S)\lambda \). [The notation in this example and the previous example are unrelated.]

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

\[ \Sigma = \begin{bmatrix} \hat{\text{var}}(\hat{S}) & \hat{\text{cov}}(\hat{S}, \hat{\lambda}) \\ \hat{\text{cov}}(\hat{\lambda}, \hat{S}) & \hat{\text{var}}(\hat{\lambda}) \end{bmatrix}. \]

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

\[ \frac{\partial \hat{S}}{\partial S} = 1, \quad \frac{\partial \hat{\lambda}}{\partial \lambda} = 0, \]

\[ \frac{\partial f}{\partial S} = -\lambda = -f'(1-S), \quad \frac{\partial f}{\partial \lambda} = 1-S. \]

Finally, the sampling variance-covariance matrix of \( S \) and \( f \) is
Thus,\
\[
\begin{bmatrix}
1 & 0 \\
-f/(1-S) & (1-S)
\end{bmatrix}
\sum
\begin{bmatrix}
1 & -f/(1-S) \\
0 & (1-S)
\end{bmatrix}.
\]

and\
\[
\hat{\text{cov}}(\hat{S}, \hat{f}) = -\hat{f}/(1-\hat{S}) \cdot \hat{\text{var}}(\hat{S}) + (1-\hat{S}) \cdot \hat{\text{cov}}(\hat{S}, \hat{\lambda})
\]

and\
\[
\hat{\text{var}}(\hat{f}) = (\hat{f}/(1-\hat{S}))^2 \cdot \hat{\text{var}}(\hat{S}) - 2\hat{f} \cdot \hat{\text{cov}}(\hat{S}, \hat{\lambda}) + (1-\hat{S})^2 \cdot \hat{\text{var}}(\hat{\lambda}).
\]