

# Appendix B

## The ‘Delta method’ ...

Suppose you have done a study, over 4 years, which yields 3 estimates of survival (say,  $\hat{\phi}_1$ ,  $\hat{\phi}_2$ , and  $\hat{\phi}_3$ ). But, suppose what you are really interested in is the estimate of the product of the three survival values (i.e., the probability of surviving from the beginning of the study to the end of the study)? While it is easy enough to derive the estimate of this product (as  $[\hat{\phi}_1\hat{\phi}_2\hat{\phi}_3]$ ), how do you estimate the variance of the estimate of the product? In other words, how do you derive an estimate of the mean, or variance, of a transformation of one or more random variables (in this case, we transform the three random variables -  $\hat{\phi}_i$  - by considering their product)?

One commonly used approach which is easily implemented, not computer-intensive, and can be robustly applied in many (but not all) situations is the so-called *Delta method* (also known as the method of propagation of errors). In this appendix, we briefly introduce the underlying background theory, and the implementation of the Delta method, to fairly typical scenarios.

### B.1. Mean and Variance of random variables

Our primary interest here is developing a method that will allow us to estimate the mean and variance for functions of random variables. Lets start by considering the formal approach for deriving these values explicitly, based on the *method of moments*. For continuous random variables, consider a continuous function  $f(x)$  on the interval  $[-\infty, +\infty]$ . The first three moments of  $f(x)$  can be written as

$$M_0 = \int_{-\infty}^{+\infty} f(x)dx$$

$$M_1 = \int_{-\infty}^{+\infty} xf(x)dx$$

$$M_2 = \int_{-\infty}^{+\infty} x^2f(x)dx$$

In the particular case that the function is a probability density (such as for a continuous random variable), then  $M_0 = 1$  (i.e., the area under the PDF must equal 1).

For example, consider the uniform distribution on the finite interval  $[a, b]$ . A uniform distribution

(sometimes also known as a rectangular distribution), is a distribution that has constant probability over the interval. The probability density function (pdf) for a continuous uniform distribution on the finite interval  $[a, b]$  is

$$P(x) = \begin{cases} 0 & \text{for } x < a \\ 1/(b-a) & \text{for } a < x < b \\ 0 & \text{for } x > b \end{cases}$$

Integrating the pdf, for  $p(x) = 1/(b-a)$ ,

$$\begin{aligned} M_0 &= \int_a^b p(x) dx \\ &= \int_a^b \frac{1}{b-a} dx = 1 \end{aligned} \quad (\text{B.1})$$

$$\begin{aligned} M_1 &= \int_a^b xp(x) dx \\ &= \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2} \end{aligned} \quad (\text{B.2})$$

$$\begin{aligned} M_2 &= \int_a^b x^2 p(x) dx \\ &= \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{3} (a^2 + ab + b^2) \end{aligned} \quad (\text{B.3})$$

We see clearly that  $M_1$  is the mean of the distribution. What about the variance? Where does the second moment  $M_2$  come in? Recall that the variance is defined as the average value of the fundamental quantity [distance from mean]<sup>2</sup>. The squaring of the distance is so the values to either side of the mean don't cancel out. Standard deviation is simply the square-root of the variance.

Given some discrete random variable  $x_i$ , with probability  $p_i$ , and mean  $\mu$ , we define the variance as

$$\text{Var} = \sum (x_i - \mu)^2 p_i$$

Note we don't have to divide by the number of values of  $x$  because the sum of the discrete probability distribution is 1 (i.e.,  $\sum p_i = 1$ ). For a continuous probability distribution, with mean  $\mu$ , we define the variance as

$$\text{Var} = \int_a^b (x - \mu)^2 p(x) dx$$

Given our moment equations, we can then write

$$\text{Var} = \int_a^b (x - \mu)^2 p(x) dx$$

$$\begin{aligned}
&= \int_a^b (x^2 - 2\mu x + \mu^2) p(x) dx \\
&= \int_a^b x^2 p(x) dx - \int_a^b 2\mu x p(x) dx + \int_a^b \mu^2 p(x) dx \\
&= \int_a^b x^2 p(x) dx - 2\mu \int_a^b x p(x) dx + \mu^2 \int_a^b p(x) dx
\end{aligned}$$

Now, if we look closely at the last line, we see that in fact the terms represent the different moments of the distribution. Thus we can write

$$\begin{aligned}
\text{Var} &= \int_a^b (x - \mu)^2 p(x) dx \\
&= \int_a^b x^2 p(x) dx - 2\mu \int_a^b x p(x) dx + \mu^2 \int_a^b p(x) dx \\
&= M_2 - 2\mu (M_1) + \mu^2 (M_0)
\end{aligned}$$

Since  $M_1 = \mu$ , and  $M_0 = 1$  then

$$\begin{aligned}
\text{Var} &= M_2 - 2\mu (M_1) + \mu^2 (M_0) \\
&= M_2 - 2\mu(\mu) + \mu^2(1) \\
&= M_2 - 2\mu^2 + \mu^2 \\
&= M_2 - \mu^2 \\
&= M_2 - (M_1)^2
\end{aligned}$$

In other words, the variance for the pdf is simply the second moment ( $M_2$ ) minus the square of the first moment ( $(M_1)^2$ ). Thus, for a continuous uniform random variable  $x$  on the interval  $[a, b]$ ,

$$\begin{aligned}
\text{Var} &= M_2 - (M_1)^2 \\
&= \frac{(a - b)^2}{12}
\end{aligned}$$

## B.2. Transformations of random variables and the Delta method

OK - thats fine. If the pdf is specified, we can use the method of moments to formally derive the mean and variance of the distribution. But, what about functions of random variables having poorly specified or unspecified distributions? Or, situations where the pdf is not easily defined?

In such cases, we may need other approaches. We'll introduce one such approach (the Delta method) here, by considering the case of a simple linear transformation of a random normal distribution.

Let

$$X_1, X_2, \dots \sim N(10, \sigma^2 = 2)$$

In other words, random deviates drawn from a normal distribution with a mean of 10, and a variance of 2. Consider some transformations of these random values. You might recall from some earlier statistics or probability class that linearly transformed normal random variables are themselves normally distributed. Consider for example,  $X_i \sim N(10, 2)$  - which we then linearly transform to  $Y_i$ , such that  $Y_i = 4X_i + 3$ .

Now, recall that for real scalar constants  $a$  and  $b$  we can show that

- i.  $E(a) = a, E(aX + b) = aE(X) + b$
- ii.  $\text{var}(a) = 0, \text{var}(aX + b) = a^2\text{var}(X)$

Thus, given  $X_i \sim N(10, 2)$  and the linear transformation  $Y_i = 4X_i + 3$ , we can write

$$Y \sim N(4(10) + 3 = 43, (4^2)2) = N(43, 32)$$

Now, an important point to note is that some transformations of the normal distribution are close to normal (i.e., are linear) and some are not. Since linear transformations of random normal values are normal, it seems reasonable to conclude that approximately linear transformations (over some range) of random normal data should also be approximately normal.

OK, to continue. Let  $X \sim N(\mu, \sigma^2)$ , and let  $Y = g(X)$ , where  $g$  is some transformation of  $X$  (in the previous example,  $g(X) = 4X + 3$ ). It is hopefully relatively intuitive that the closer  $g(X)$  is to linear over the likely range of  $X$  (i.e., within 3 or so standard deviations of  $\mu$ ), the closer  $Y = g(X)$  will be to normally distributed. From calculus, we recall that if you look at any differentiable function over a narrow enough region, the function appears approximately linear. The approximating line is the tangent line to the curve, and its slope is the derivative of the function.

Since most of the mass (i.e., most of the random values) of  $X$  is concentrated around  $\mu$ , let's figure out the tangent line at  $\mu$ , using two different methods. First, we know that the tangent line passes through  $(\mu, g(\mu))$ , and that its slope is  $g'(\mu)$  (we use the ' $g'$ ' notation to indicate the first derivative of the function  $g$ ). Thus, the equation of the tangent line is  $Y = g'(\mu)X + b$  for some  $b$ . Replacing  $(X, Y)$  with the known point  $(\mu, g(\mu))$ , we find  $g(\mu) = g'(\mu)\mu + b$  and so  $b = g(\mu) - g'(\mu)\mu$ . Thus, the equation of the tangent line is  $Y = g'(\mu)X + g(\mu) - g'(\mu)\mu$ .

Now for the big step - we can derive an approximation to the same tangent line by using a *Taylor series expansion* of  $g(x)$  (to first order) around  $X = \mu$ :

$$\begin{aligned} Y &= g(X) \\ &\approx g(\mu) + g'(\mu)(X - \mu) + \epsilon \\ &= g'(\mu)X + g(\mu) - g'(\mu)\mu + \epsilon \end{aligned}$$

OK, at this point you might be asking yourself 'so what?'. Well, suppose that  $X \sim N(\mu, \sigma^2)$  and  $Y = g(X)$ , where  $g'(\mu) \neq 0$ . Then, whenever the tangent line (derived earlier) is approximately correct over the likely range of  $X$  (i.e., if the transformed function is approximately linear over the likely range of  $X$ ), then the transformation  $Y = g(X)$  will have an approximate normal distribution. That approximate normal distribution may be found using the usual rules for linear transformations

of normals.

Thus, to first order,

$$E(Y) = g'(\mu)\mu + g(\mu) - g'(\mu)\mu = g(\mu)$$

$$\begin{aligned}\text{var}(Y) &= \text{var}(g(X)) = (g(X) - g(\mu))^2 \\ &= (g'(\mu)(X - \mu))^2 \\ &= (g'(\mu))^2(X - \mu)^2 \\ &= (g'(\mu))^2\text{var}(X)\end{aligned}$$

These first-order approximations are usually referred to as the *Delta method*.

begin sidebar

### Taylor series expansions?

A very important, and frequently used tool. If you have no familiarity at all with series expansions, here is a (very) short introduction. Briefly, the *Taylor series* is a power series expansion of an infinitely differentiable real (or complex) function defined on an open interval around some specified point. For example, a one-dimensional Taylor series is an expansion of a real function  $f(x)$  about a point  $x = a$  over the interval  $(a - r, a + r)$ , is given as:

$$f(x) \approx f(a) + \frac{f'(a)(x-a)}{1!} + \frac{f''(a)(x-a)^2}{2!} + \dots$$

where  $f'(a)$  is the first derivative of  $f$  with respect to  $a$ ,  $f''(a)$  is the second derivative of  $f$  with respect to  $a$ , and so on.

For example, suppose the function is  $f(x) = e^x$ . The convenient fact about this function is that all its derivatives are equal to  $e^x$  as well (i.e.,  $f(x) = e^x, f'(x) = e^x, f''(x) = e^x, \dots$ ). In particular,  $f^{(n)}(x) = e^x$  so that  $f^{(n)}(0) = 1$ . This means that the coefficients of the Taylor series are given by

$$a_n = \frac{f^{(n)}(0)}{n!} = \frac{1}{n!}$$

and so the Taylor series is given by

$$1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \dots + \frac{x^n}{n!} + \dots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

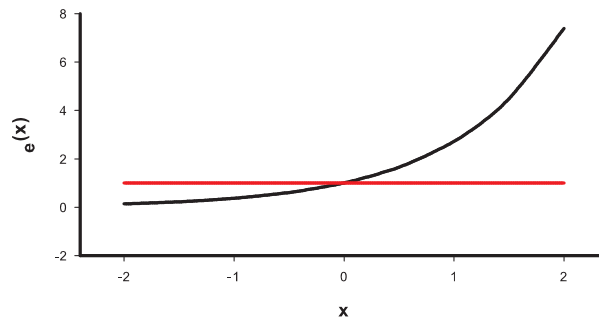
The primary utility of such a power series in simple application is that differentiation and integration of power series can be performed term by term and is hence particularly (or, at least relatively) easy. In addition, the (truncated) series can be used to compute function values approximately.

Now, let's look at an example of the "fit" of a Taylor series to a familiar function, given a certain number of terms in the series. For our example, we'll expand the function  $f(x) = e^x$ , at  $x = a = 0$ , on the interval  $(a - 2, a + 2)$ , for  $n = 0, n = 1, n = 2, \dots$  (where  $n$  is the number of terms in the series).

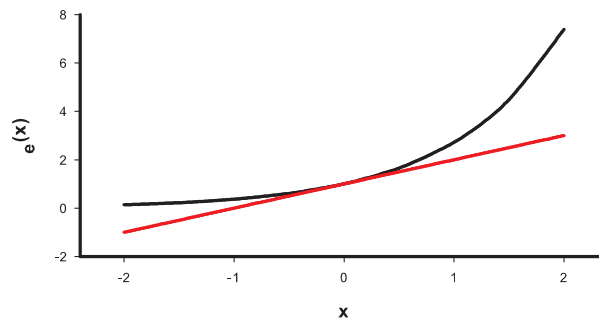
For  $n = 0$ , the Taylor expansion is a scalar constant (1):

$$f(x) \approx 1$$

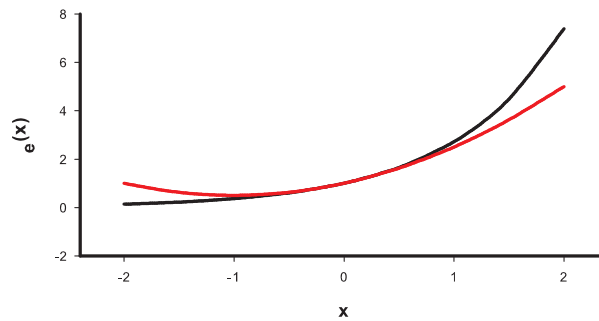
which is obviously a poor approximation to the function  $f(x) = e^x$  at any point. This is shown clearly in the figure at the top of the next page - the black line in the figure is the function  $f(x) = e^x$ , evaluated over the interval  $(-2, 2)$ , and the red line is the Taylor series approximation for  $n = a = 0$ .



What happens when we add higher order terms? Here is the plot of the Taylor series for  $n = 1$ .



Hmmm...a bit better. What about  $n = 2$ ?



Now we're getting somewhere. We see that when we add more terms (i.e., use a higher-order series), the fit gets progressively better. Often, for 'nice, smooth' functions (i.e., those nearly linear at the point of interest), we don't need many terms at all. For this example,  $n = 4$  yields a near-perfect fit (over the interval  $(-2, 2)$ ).

---

end sidebar

---

### B.3. Transformations of one variable

OK, enough background for now. Lets see some applications. Lets check the Delta method out in a case where we know the answer. Assume we have an estimate of density  $\widehat{D}$  and its conditional sampling variance,  $\widehat{\text{var}}(D_s)$ . We want to multiply this by some constant  $c$  to make it comparable with other values from the literature. Thus, we want  $\widehat{D}_s = g(D) = c\widehat{D}$  and  $\widehat{\text{var}}D_s$ .

The Delta method gives

$$\begin{aligned}\widehat{\text{var}}(D_s) &= (g'(D))^2 \hat{\sigma}_D^2 \\ &= \left( \frac{\partial \widehat{D}_s}{\partial \widehat{D}} \right)^2 \cdot \widehat{\text{var}}(D) \\ &= c^2 \cdot \widehat{\text{var}}(\widehat{D})\end{aligned}$$

which we know to be true for the variance of a random variable multiplied by a real constant.

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

Still another example - you have some parameter  $\theta$ , which you transform by dividing it by some constant  $c$ . Thus, by the Delta method,

$$\widehat{\text{var}}\left(\frac{\widehat{\theta}}{c}\right) = \left(\frac{1}{c}\right)^2 \cdot \widehat{\text{var}}(\widehat{\theta})$$

#### B.3.1. A potential complication - violation of assumptions

A final - and important - example for transformations of single variables. The importance lies in the demonstration that the Delta method does not always work - remember, it assumes that the transformation is approximately linear over the expected range of the parameter. Suppose one has an MLE for the mean and estimated variance for some parameter  $\theta$  which is bounded random uniform on the interval  $[0, 2]$ . Suppose you want to transform this parameter such that

$$\psi = e^{(\theta)}$$

(Recall that this is a convenient transformation since the derivative of  $e^x$  is  $e^x$ , making the calculations very simple). Now, based on the Delta method, the variance for  $\psi$  would be estimated as

$$\begin{aligned}\widehat{\text{var}}(\psi) &= \left( \frac{\partial \widehat{\psi}}{\partial \widehat{\theta}} \right)^2 \cdot \widehat{\text{var}}(\widehat{\theta}) \\ &= (e^{\widehat{\theta}})^2 \cdot \widehat{\text{var}}(\widehat{\theta})\end{aligned}$$

Now, suppose that  $\widehat{\theta} = 1.0$ , and  $\widehat{\text{var}}(\widehat{\theta}) = 0.33\dot{3}$ . Then, from the Delta method,

$$\begin{aligned}\widehat{\text{var}}(\widehat{\psi}) &= \left(e^{\theta}\right)^2 \cdot \widehat{\text{var}}(\widehat{\theta}) \\ &= (7.38906)(0.3\ddot{3}) \\ &= 2.46302\end{aligned}$$

OK, so what's the problem? Well, let's derive the variance of  $\psi$  using the method of moments. To do this, we need to integrate the pdf (uniform, in this case) over some range. Since the variance of a uniform distribution is  $(b-a)^2/12$ , and if  $b$  and  $a$  are symmetric around the mean (1.0), then we can show by algebra that given a variance of 0.33, then  $a = 0$  and  $b = 2$ .

Given a uniform distribution, the pdf is  $p(\theta) = 1/(b-a)$ . Thus, by the method of moments,

$$\begin{aligned}M_1 &= \int_a^b g(x)p(x)dx \\ &= \int_a^b \frac{g(x)}{b-a}dx \\ &= -\frac{e^b - e^a}{a-b}\end{aligned}$$

$$\begin{aligned}M_2 &= \int_a^b \frac{g(x)^2}{b-a}dx \\ &= \frac{1}{2} \frac{e^{2a} - e^{2b}}{a-b}\end{aligned}$$

Thus, by moments,  $\widehat{\text{var}}(E(\psi))$  is

$$\begin{aligned}\widehat{\text{var}}(E(\psi)) &= M_2 - (M_1)^2 \\ &= \frac{1}{2} \frac{e^{2b} - e^{2a}}{-b+a} - \frac{(e^b - e^a)^2}{(a-b)^2}\end{aligned}$$

If  $a = 0$  and  $b = 2$ , then

$$\widehat{\text{var}}(E(\psi)) = M_2 - (M_1)^2 = \frac{1}{2} \frac{e^{2b} - e^{2a}}{-b+a} - \frac{(e^b - e^a)^2}{(a-b)^2} = 3.19453$$

which is not particularly close to the value estimated by the Delta method (2.46302).

Why the discrepancy? As discussed earlier, the Delta method rests on the assumption the first-order Taylor expansion around the parameter value is effectively linear over the range of values likely to be encountered. Since in this example we're using a uniform pdf, then all values between  $a$  and  $b$  are equally likely. Thus, we might anticipate that as the interval between  $a$  and  $b$  gets smaller, then the approximation to the variance (which will clearly decrease) will get better and better (since the

smaller the interval, the more likely it is that the function is approximately linear over that range). For example, if  $a = 0.5$  and  $b = 1.5$  (same mean of 1.0), then the true variance of  $\theta$  will be 0.083. Thus, by the Delta method, the estimated variance of  $\psi$  will be 0.61575, while by the method of moments (which is exact), the variance will be 0.65792. Clearly, the proportional difference between the two values has declined markedly. But, we achieved this 'improvement' by artificially reducing the true variance of the untransformed variable  $\theta$ . Obviously, we can't do this in reality.

So, what are the options? Well, one immediate option is to use a higher order Taylor series approximation - by including higher order terms, we can achieve a better 'fit' to the function (see the preceding sidebar). If we used a second-order TSE,

$$E(g(x)) \approx g(\mu) + \frac{1}{2}g''(\mu)\sigma^2 \quad (\text{B.4})$$

$$\text{Var}(g(x)) \approx g'(\mu)^2\sigma^2 + \frac{1}{4}(g''(\mu))^2(\text{Var}(x^2) - 4\mu^2\sigma^2) \quad (\text{B.5})$$

we should do a bit better. For the variance estimate, we need to know  $\text{var}(x^2)$ , which for a continuous uniform distribution by the method of moments is

$$1/5 \frac{b^5 - a^5}{b - a} - 1/9 \frac{(b^3 - a^3)^2}{(b - a)^2}$$

Thus, from the second-order approximation, and again assuming  $a = 0$  and  $b = 2$ , then  $\widehat{\text{var}}(\psi)$  is

$$\begin{aligned} \widehat{\text{var}}(\psi) &= (e^\theta)^2 \cdot \text{var}(\theta) + \frac{1}{4}(e^\theta)^2 \cdot \text{var}(\theta^2) - 4\mu^2\text{var}(\theta) \\ &= 3.756316 \end{aligned}$$

which is closer (proportionately) to the true variance (3.19453) than was the estimate using only the first-order TSE (2.46302). The reason that even a second-order approximation isn't 'much closer' is because the transformation is very non-linear over the range of the data (uniform  $[0, 2]$  in this case), such that the second order approximation doesn't 'fit' particularly well over this range - this is shown graphically at the top of page B-7.

So, we see that the classical Delta method - which is based on a first-order Taylor series expansion of the transformed function - may not do particularly well if the function is highly non-linear over the range of values being examined.

Of course, it would be fair to note that the preceding example made the assumption that the distribution was random uniform over the interval. For most of our work with **MARK**, the interval is likely to have asymmetric mass around the estimate, typically  $\beta$ . As such, most of data, and thus the transformed data, will actually fall closer to the parameter value in question (the mean in this example) than we've demonstrated here. So much so, that the discrepancy between the first order 'Delta' approximation to the variance and the true value of the variance will likely be significantly smaller than shown here, even for a strongly non-linear transformation. We leave it to you as an exercise to prove this for yourself. But, this point notwithstanding, it is important to be aware of the assumptions underlying the Delta method - if your transformation is non-linear, and there is considerable variation in your data, the first-order approximation may not be particular good. Fortunately, use of second order Taylor series approximations is not heroically difficult (the challenge

is usually coming up with  $\widehat{\text{var}(X^2)}$ . If the pdf for the untransformed data is specified (which is essentially equivalent to assuming an informative prior), then you can derive  $\widehat{\text{var}(X^2)}$  fairly easily using the method of moments.

#### B.4. Transformations of two or more variables

Clearly, we are often interested in transformations involving more than one variable. Fortunately, there are also multivariate generalizations of the Delta method.

Suppose you've estimated  $p$  different random variables  $X_1, X_2, \dots, X_p$ . In matrix notation, these variables would configure a  $p \times 1$  random vector

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$$

which has a mean vector

$$\boldsymbol{\mu} = \begin{pmatrix} EX_1 \\ EX_2 \\ \vdots \\ EX_p \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}$$

and the  $p \times p$  variance-covariance matrix is

$$\begin{pmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) & \dots & \text{cov}(X_1, X_p) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) & \dots & \text{cov}(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_p, X_1) & \text{cov}(X_p, X_2) & \dots & \text{var}(X_p) \end{pmatrix}$$

Note that if the variables are independent, then the off-diagonal elements (i.e., the covariance terms) are all zero.

Then, for a  $k \times p$  matrix of constants  $\mathbf{A} = a_{ij}$ , the expectation of a random vector  $\mathbf{Y} = \mathbf{A}\mathbf{X}$  is given as

$$\begin{pmatrix} EY_1 \\ EY_2 \\ \vdots \\ EY_p \end{pmatrix} = \mathbf{A}\boldsymbol{\mu}$$

with a variance-covariance matrix

$$\text{cov}(\mathbf{Y}) = \mathbf{A}\Sigma\mathbf{A}^T$$

Now, using the same logic we first considered for developing the Delta method for a single variable, for each  $x_i$  near  $\mu_i$ , we can write

$$\mathbf{y} = \begin{pmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_p(x) \end{pmatrix} \approx \begin{pmatrix} g_1(\mu) \\ g_2(\mu) \\ \vdots \\ g_p(\mu) \end{pmatrix} + \mathbf{D}(x - \mu)$$

where  $\mathbf{D}$  is the matrix of partial derivatives of  $g_i$  with respect to  $x_j$ , evaluated at  $(x - \mu)$ .

Now, as with the single-variable Delta method, if the variances of the  $X_i$  are small (so that with high probability  $Y$  is near  $\mu$ , such that the linear approximation is usually valid), then to first-order

$$\begin{pmatrix} EY_1 \\ EY_2 \\ \vdots \\ EY_p \end{pmatrix} = \begin{pmatrix} g_1(\mu) \\ g_2(\mu) \\ \vdots \\ g_p(\mu) \end{pmatrix} \quad \text{var}(\mathbf{Y}) = \mathbf{D}\Sigma\mathbf{D}^T$$

#### **example (1) - variance of product of survival rates**

Let's consider the application of the Delta method in estimating sampling variances of a fairly common function - the product of several parameter estimates.

Now, from the preceding, we see that

$$\text{var}(\mathbf{Y}) = \mathbf{D}\Sigma\mathbf{D}^T$$

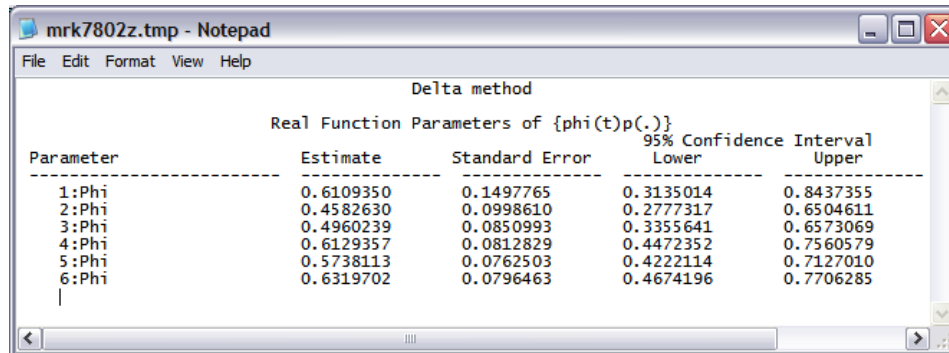
which we write out more fully as

$$\begin{aligned} \text{var}(\mathbf{Y}) &= \mathbf{D}\Sigma\mathbf{D}^T \\ &= \left( \frac{\partial(\hat{Y})}{\partial(\hat{\theta})} \right) \cdot \hat{\Sigma} \cdot \left( \frac{\partial(\hat{Y})}{\partial(\hat{\theta})} \right)^T \end{aligned}$$

where  $Y$  is some linear or nonlinear function of the parameter estimates  $\hat{\theta}_1, \hat{\theta}_2, \dots$ . The first term on the RHS of the variance expression is a row vector containing partial derivatives of  $Y$  with respect to each of these parameters ( $\hat{\theta}_1, \hat{\theta}_2, \dots$ ). The right-most term of the RHS of the variance expression is simply a transpose of this row vector (i.e., a column vector). The middle-term is simply the estimated variance-covariance matrix for the parameters.

OK, let's try an example - let's use estimates from the male European dipper data set (yes, again). We'll fit model  $\{\phi_i p_i\}$  to these data. Recall that there are 7 occasions in this particular data set, generating 6 estimates of survival for a time-dependent model (since we've constrained  $p$  to be

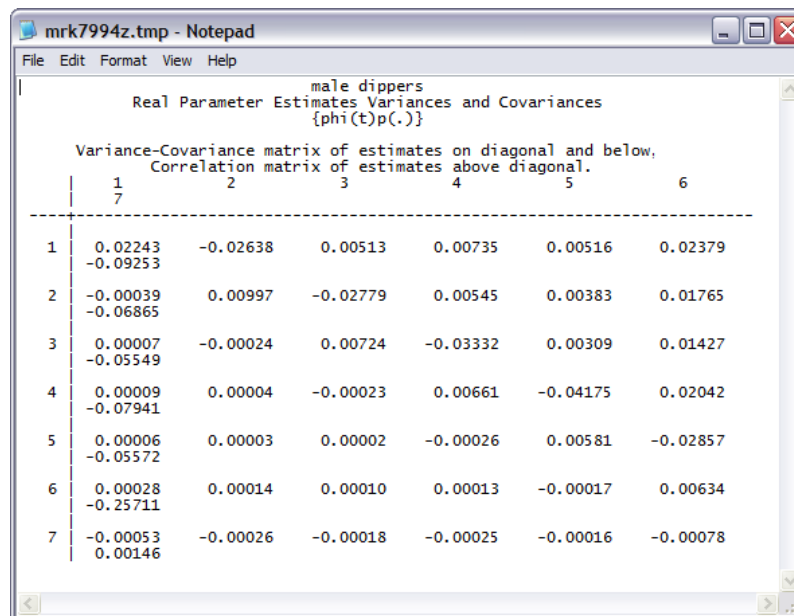
constant over time, all 6 estimates are separately identifiable). Here are the estimated survival rates from this model:



Delta method				
Real Function Parameters of {phi(t)p(.)}				
Parameter	Estimate	Standard Error	95% Confidence Interval Lower	95% Confidence Interval Upper
1:Phi	0.6109350	0.1497765	0.3135014	0.8437355
2:Phi	0.4582630	0.0998610	0.2777317	0.6504611
3:Phi	0.4960239	0.0850993	0.3355641	0.6573069
4:Phi	0.6129357	0.0812829	0.4472352	0.7560579
5:Phi	0.5738113	0.0762503	0.4222114	0.7127010
6:Phi	0.6319702	0.0796463	0.4674196	0.7706285

Suppose we're interested in the probability of surviving from the start of the first interval to the end of the third interval. Well, the point-estimate of this probability is easy enough - its simply  $(\hat{\phi}_1 \times \hat{\phi}_2 \times \hat{\phi}_3)$ , which, from the table of results shown on the preceding page, and assuming these estimates come from an adequate model, is  $(0.6109350 \times 0.4582630 \times 0.4960239) = 0.138871$ . So, the probability of a male Dipper surviving over the first three intervals is  $\sim 14\%$  (again, assuming that our time-dependent survival model is a valid model).

To derive the estimate of the variance of the product, we will also need the **variance-covariance matrix** for the survival estimates. You can generate the matrix easily in **MARK** by selecting 'Output-Specific Model Output-Variance-Covariance Matrices-Real Estimates'. Here they are for the male Dipper data, for model  $\{\phi_t p.\}$ :



male dippers						
Real Parameter Estimates Variances and Covariances						
{phi(t)p(.)}						
Variance-Covariance matrix of estimates on diagonal and below.						
Correlation matrix of estimates above diagonal.						
	1	2	3	4	5	6
1	0.02243 -0.09253	-0.02638	0.00513	0.00735	0.00516	0.02379
2	-0.00039 -0.06865	0.00997	-0.02779	0.00545	0.00383	0.01765
3	0.00007 -0.05549	-0.00024	0.00724	-0.03332	0.00309	0.01427
4	0.00009 -0.07941	0.00004	-0.00023	0.00661	-0.04175	0.02042
5	0.00006 -0.05572	0.00003	0.00002	-0.00026	0.00581	-0.02857
6	0.00028 -0.25711	0.00014	0.00010	0.00013	-0.00017	0.00634
7	-0.00053 0.00146	-0.00026	-0.00018	-0.00025	-0.00016	-0.00078

In the Notepad output from **MARK**, the variance-covariance values are *below* the diagonal, whereas the standardized correlation values are *above* the diagonal.

However, it is **very important** to note that the V-C matrix that **MARK** outputs to the Notepad is *rounded* to 5 significant digits. For the actual calculations, we need to use the full precision values. To get those, you need to either (i) output the V-C matrix into a dBase file (which you could then open with dBase, or Excel), or (ii) copy the V-C matrix into the Windows clipboard, and then paste it into some other application. Failure to use the full precision V-C matrix will often (almost always, in fact) lead to ‘rounding errors’. Here is the ‘full precision’ V-C matrix for the survival values:

$$\begin{aligned} \widehat{\text{cov}}(Y) &= \begin{pmatrix} \text{var}(\phi_1) & \text{cov}(\phi_1, \phi_2) & \text{cov}(\phi_1, \phi_3) \\ \text{cov}(\phi_2, \phi_1) & \text{var}(\phi_2) & \text{cov}(\phi_2, \phi_3) \\ \text{cov}(\phi_3, \phi_1) & \text{cov}(\phi_3, \phi_2) & \text{var}(\phi_3) \end{pmatrix} \\ &= \begin{pmatrix} 0.0224330125 & -0.0003945405 & 0.0000654469 \\ -0.0003945405 & 0.0099722201 & -0.0002361998 \\ 0.0000654469 & -0.0002361998 & 0.0072418858 \end{pmatrix} \end{aligned}$$

Now what? First, we need to identify the transformation we’re applying to our estimates ( $\hat{\phi}_1$ ,  $\hat{\phi}_2$ , and  $\hat{\phi}_3$ ). In this case, the transformation (which we’ll call  $Y$ ) is simple - its the product of the three estimated survival rates. Conveniently, this make differentiating the transformation straightforward.

So, here is the variance estimator, in full:

$$\widehat{\text{var}}(Y) = \begin{bmatrix} \left(\frac{\partial(\hat{Y})}{\partial\hat{\phi}_1}\right) & \left(\frac{\partial(\hat{Y})}{\partial\hat{\phi}_2}\right) & \left(\frac{\partial(\hat{Y})}{\partial\hat{\phi}_3}\right) \end{bmatrix} \cdot \widehat{\Sigma} \cdot \begin{bmatrix} \left(\frac{\partial(\hat{Y})}{\partial\hat{\phi}_1}\right) \\ \left(\frac{\partial(\hat{Y})}{\partial\hat{\phi}_2}\right) \\ \left(\frac{\partial(\hat{Y})}{\partial\hat{\phi}_3}\right) \end{bmatrix}$$

Now, each of the partial derivatives is easy enough. For example, since  $\hat{Y} = \hat{\phi}_1\hat{\phi}_2\hat{\phi}_3$ , then  $\partial\hat{Y}/\partial\hat{\phi}_1 = \hat{\phi}_2\hat{\phi}_3$ . And so on. So,

$$\begin{aligned} \widehat{\text{var}}(Y) &= \begin{bmatrix} \left(\frac{\partial(\hat{Y})}{\partial\hat{\phi}_1}\right) & \left(\frac{\partial(\hat{Y})}{\partial\hat{\phi}_2}\right) & \left(\frac{\partial(\hat{Y})}{\partial\hat{\phi}_3}\right) \end{bmatrix} \cdot \widehat{\Sigma} \cdot \begin{bmatrix} \left(\frac{\partial(\hat{Y})}{\partial\hat{\phi}_1}\right) \\ \left(\frac{\partial(\hat{Y})}{\partial\hat{\phi}_2}\right) \\ \left(\frac{\partial(\hat{Y})}{\partial\hat{\phi}_3}\right) \end{bmatrix} \\ &= \begin{bmatrix} (\hat{\phi}_2\hat{\phi}_3) & (\hat{\phi}_1\hat{\phi}_3) & (\hat{\phi}_1\hat{\phi}_2) \end{bmatrix} \cdot \widehat{\Sigma} \cdot \begin{bmatrix} (\hat{\phi}_2\hat{\phi}_3) \\ (\hat{\phi}_1\hat{\phi}_3) \\ (\hat{\phi}_1\hat{\phi}_2) \end{bmatrix} \end{aligned}$$

OK, what about the variance-covariance matrix? Well, from the preceding we see that

$$\begin{aligned}\widehat{\text{cov}}(\hat{Y}) &= \begin{pmatrix} \text{var}(\phi_1) & \text{cov}(\phi_1, \phi_2) & \text{cov}(\phi_1, \phi_3) \\ \text{cov}(\phi_1, \phi_1) & \text{var}(\phi_2) & \text{cov}(\phi_2, \phi_3) \\ \text{cov}(\phi_3, \phi_1) & \text{cov}(\phi_3, \phi_2) & \text{var}(\phi_3) \end{pmatrix} \\ &= \begin{pmatrix} 0.0224330125 & -0.0003945405 & 0.0000654469 \\ -0.0003945405 & 0.0099722201 & -0.0002361998 \\ 0.0000654469 & -0.0002361998 & 0.0072418858 \end{pmatrix}\end{aligned}$$

So, after brushing up on your linear algebra skills, we see that

$$\begin{aligned}\widehat{\text{var}}(\hat{Y}) &= \begin{bmatrix} (\hat{\phi}_2\hat{\phi}_3) & (\hat{\phi}_1\hat{\phi}_3) & (\hat{\phi}_1\hat{\phi}_2) \end{bmatrix} \cdot \hat{\Sigma} \cdot \begin{bmatrix} (\hat{\phi}_2\hat{\phi}_3) \\ (\hat{\phi}_1\hat{\phi}_3) \\ (\hat{\phi}_1\hat{\phi}_2) \end{bmatrix} \\ &= \begin{bmatrix} (\hat{\phi}_2\hat{\phi}_3) & (\hat{\phi}_1\hat{\phi}_3) & (\hat{\phi}_1\hat{\phi}_2) \end{bmatrix} \cdot \begin{pmatrix} \text{var}(\phi_1) & \text{cov}(\phi_1, \phi_2) & \text{cov}(\phi_1, \phi_3) \\ \text{cov}(\phi_1, \phi_1) & \text{var}(\phi_2) & \text{cov}(\phi_2, \phi_3) \\ \text{cov}(\phi_3, \phi_1) & \text{cov}(\phi_3, \phi_2) & \text{var}(\phi_3) \end{pmatrix} \cdot \begin{bmatrix} (\hat{\phi}_2\hat{\phi}_3) \\ (\hat{\phi}_1\hat{\phi}_3) \\ (\hat{\phi}_1\hat{\phi}_2) \end{bmatrix}\end{aligned}$$

Clearly, this expression is getting more and more ‘impressive’ as we progress. Well, here is the resulting expression (written in piecewise fashion to make it easier to see the basic pattern):

$$\begin{aligned}\widehat{\text{var}}(\hat{Y}) &\cong \hat{\phi}_2^2\hat{\phi}_3^2(\widehat{\text{var}}_1) \\ &\quad + 2\hat{\phi}_2\hat{\phi}_3^2\hat{\phi}_1(\widehat{\text{cov}}_{1,2}) \\ &\quad + 2\hat{\phi}_2^2\hat{\phi}_3\hat{\phi}_1(\widehat{\text{cov}}_{1,3}) \\ &\quad + \hat{\phi}_1^2\hat{\phi}_3^2(\widehat{\text{var}}_2) \\ &\quad + 2\hat{\phi}_1^2\hat{\phi}_3\hat{\phi}_2(\widehat{\text{cov}}_{2,3}) \\ &\quad + \hat{\phi}_1^2\hat{\phi}_2^2(\widehat{\text{var}}_3)\end{aligned}$$

Whew - a lot of work (and if you think this equation looks ‘impressive’, try it using a second-order Taylor series approximation!). But, under some assumptions, the Delta method does rather well in allowing you to derive an estimate of the sampling variance for functions of random variables (or, as we’ve described, functions of estimated parameters). So, after substituting in our estimates for  $\phi_i$  and the variances and covariances, our estimate for the sampling variance of the product  $\hat{Y} = (\hat{\phi}_1\hat{\phi}_2\hat{\phi}_3)$  is (approximately) 0.0025565.

**example (2) - variance of estimate 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 some amount of money (say, \$50) will be given to people reporting these special tags. It is assumed that \$50 is sufficient to ensure 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 as  $\lambda$ . Thus,

$$\hat{\lambda} = \frac{\hat{f}_c}{\hat{f}_r}$$

Now, note that both recovery rates are binomials. Thus,

$$\widehat{\text{var}}(f_c) = \frac{\hat{f}_c(1-\hat{f}_c)}{N_c} \quad \widehat{\text{var}}(f_r) = \frac{\hat{f}_r(1-\hat{f}_r)}{N_r}$$

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

$$\begin{pmatrix} \widehat{\text{var}}(f_c) & 0 \\ 0 & \widehat{\text{var}}(f_r) \end{pmatrix}$$

Next, we need the derivatives of  $\lambda$  with respect to  $f_c$  and  $f_r$ :

$$\frac{\partial \hat{\lambda}}{\partial \hat{f}_c} = \frac{1}{\hat{f}_r} \quad \frac{\partial \hat{\lambda}}{\partial \hat{f}_r} = -\frac{\hat{f}_c}{\hat{f}_r^2}$$

Thus,

$$\widehat{\text{var}}(\lambda) = \begin{pmatrix} \frac{1}{\hat{f}_r} & -\frac{\hat{f}_c}{\hat{f}_r^2} \end{pmatrix} \begin{pmatrix} \widehat{\text{var}}(f_c) & 0 \\ 0 & \widehat{\text{var}}(f_r) \end{pmatrix} \begin{pmatrix} \frac{1}{\hat{f}_r} \\ -\frac{\hat{f}_c}{\hat{f}_r^2} \end{pmatrix}$$

**example (3) - variance and SE of back-transformed estimates - simple**

The basic idea behind this final worked example was introduced back in Chapter 6 - in that chapter, we demonstrate how we can ‘back-transform’ from the estimate of  $\beta$  on the logit scale to an estimate

of some parameter  $\theta$  (e.g.,  $\phi$  or  $p$ ) on the probability scale (which is bounded  $[0, 1]$ ). But, we're clearly also interested in an estimate of the variance (precision) of our estimate, on both scales. Your first thought might be to simply back-transform from the link function (in our example, the logit link), to the probability scale, just as we did above. But, as discussed in chapter 6, this does not work.

For example, consider the male Dipper data. Using the logit link, we fit model  $\{\phi, p.\}$  to the data - no time-dependence for either parameter. Lets consider only the estimate for  $\hat{\phi}$ . The estimate for  $\beta$  for  $\phi$  is 0.2648275. Thus, our estimate of  $\hat{\phi}$  on the probability scale is

$$\begin{aligned}\hat{\phi} &= \frac{e^{0.2648275}}{1 + e^{0.2648275}} \\ &= \frac{1.303206}{2.303206} = 0.5658226\end{aligned}$$

which is exactly what **MARK** reports (to within rounding error).

But, what about the variance? Well, if we look at the  $\beta$  estimates, **MARK** reports that the standard error for the estimate of  $\beta$  corresponding to survival is 0.1446688. If we simply back-transform this from the logit scale to the probability scale, we get

$$\begin{aligned}\widehat{\text{SE}} &= \frac{e^{0.1446688}}{1 + e^{0.1446688}} \\ &= \frac{1.155657}{2.155657} = 0.5361043\end{aligned}$$

However, **MARK** reports the estimated standard error for  $\phi$  as 0.0355404, which isn't even remotely close to our back-transformed value of 0.5361043.

What has happened? Well, hopefully you now realize that you're 'transforming' the estimate from one scale (logit) to another (probability). And, since you're working with a 'transformation', you need to use the Delta method to estimate the variance of the back-transformed parameter. Since

$$\hat{\phi} = \frac{e^{\hat{\beta}}}{1 + e^{\hat{\beta}}}$$

then

$$\begin{aligned}\widehat{\text{var}}(\hat{\phi}) &= \left( \frac{\partial \hat{\phi}}{\partial \hat{\beta}} \right)^2 \times \widehat{\text{var}}(\hat{\beta}) \\ &= \left( \frac{e^{\hat{\beta}}}{1 + e^{\hat{\beta}}} - \frac{(e^{\hat{\beta}})^2}{1 + (e^{\hat{\beta}})^2} \right)^2 \times \widehat{\text{var}}(\hat{\beta})\end{aligned}$$

$$= \left( \frac{e^{\hat{\beta}}}{(1 + e^{\hat{\beta}})^2} \right)^2 \times \widehat{\text{var}}(\hat{\beta})$$

It is worth noting that if

$$\hat{\phi} = \frac{e^{\hat{\beta}}}{1 + e^{\hat{\beta}}}$$

then it can be easily shown that

$$\hat{\phi}(1 - \hat{\phi}) = \frac{e^{\hat{\beta}}}{(1 + e^{\hat{\beta}})^2}$$

which is the derivative of  $\phi$  with respect to  $\beta$ . So, we could rewrite our expression for the variance of  $\hat{\phi}$  conveniently as

$$\begin{aligned} \widehat{\text{var}}(\hat{\phi}) &= \left( \frac{e^{\hat{\beta}}}{(1 + e^{\hat{\beta}})^2} \right)^2 \times \widehat{\text{var}}(\hat{\beta}) \\ &= (\hat{\phi}(1 - \hat{\phi}))^2 \times \widehat{\text{var}}(\hat{\beta}) \end{aligned}$$

From **MARK**, the estimate of the SE for  $\hat{\beta}$  was 0.1446688. Thus, the estimate of  $\widehat{\text{var}}(\hat{\beta})$  is  $0.1446688^2 = 0.02092906$ . Given the estimate of  $\hat{\beta}$  of 0.2648275, we substitute into the preceding expression, which yields

$$\begin{aligned} \widehat{\text{var}}(\hat{\phi}) &= \left( \frac{e^{\hat{\beta}}}{(1 + e^{\hat{\beta}})^2} \right)^2 \times \widehat{\text{var}}(\hat{\beta}) \\ &= 0.0603525 \times 0.02092906 = 0.001263 \end{aligned}$$

So, the estimated SE for  $\hat{\phi}$  is  $\sqrt{0.001263} = 0.0355404$ , which is what is reported by **MARK** (again, within rounding error).

begin sidebar

#### SE and 95% CI

The standard approach to calculating 95% confidence limits for some parameter  $\theta$  is  $\theta \pm (1.96 \times \text{SE})$ . Is this how **MARK** calculates the 95% CI on the real probability scale? Well, take the example we just considered - the estimated SE for  $\hat{\phi} = 0.5658226$  was  $\sqrt{0.001263} = 0.0355404$ . So, you might assume that the 95% CI on the real probability scale would be  $0.5658226 \pm (2 \times 0.0355404) = [0.4947418, 0.6369034]$ .

However, this is not what is reported by **MARK** -  $[0.4953193, 0.6337593]$ , which is quite close, but not exactly the same. Why the difference? The difference is because **MARK** first calculated the 95% CI on the logit scale,

before back-transforming to the real probability scale. So, for our estimate of  $\hat{\phi}$ , the 95% CI on the logit scale for  $\hat{\beta} = 0.2648275$  is  $[-0.0187234, 0.5483785]$ , which, when back-transformed to the real probability scale is  $[0.4953193, 0.6337593]$ , which is what is reported by **MARK**. In this case, the very small difference between the two CI's is because the parameter estimate was quite close to 0.5. In such cases, not only will the 95% CI be nearly the same (for estimates of 0.5, it will be identical), but they will also be symmetrical.

However, because the logit transform is not linear, the *reconstituted* 95% CI will not be symmetrical around the parameter estimate, especially for parameters estimated near the  $[0, 1]$  boundaries. For example, consider the estimate for  $\hat{p} = 0.9231757$ . On the logit scale, the 95% CI for the  $\beta$  corresponding to  $p$  (SE=0.5120845) is  $[1.4826128, 3.4899840]$ . The back-transformed CI is  $[0.8149669, 0.9704014]$ . This CI is clearly **not** symmetric around  $\hat{p} = 0.9231757$ . Essentially the degree of asymmetry is a function of how close the estimated parameter is to either the 0 or 1 boundary. Further, the estimated variance for  $\hat{p}$

$$\begin{aligned}\widehat{\text{var}}(\hat{p}) &= (\hat{p}(1 - \hat{p}))^2 \times \widehat{\text{var}}(\hat{\beta}) \\ &= (0.9231757(1 - 0.9231757))^2 \times 0.262231 = 0.001319\end{aligned}$$

yields an estimated SE of 0.036318 on the normal probability scale (which is what is reported by **MARK**). Estimating the 95% CI on the normal probability scale simply as  $0.9231757 \pm (2 \times 0.036318)$  yields  $[0.85054, 0.99581]$ , which is clearly quite a bit different, and more symmetrical, than what is reported by **MARK** (from above,  $[0.8149669, 0.9704014]$ ).

**MARK** uses the back-transformed CI to ensure that the reported CI is bounded  $[0, 1]$ . As the estimated parameter approaches either the 0 or 1 boundary, the degree of asymmetry in the back-transformed 95% CI that **MARK** reports will increase.

---

end sidebar

---

Got it? Well, as a final test, consider the following, more difficult, example of back-transforming the CI from a model fit using individual covariates (discussed in Chapter 11).

#### **example (4) - variance and SE of back-transformed estimates - harder**

Recall that in that chapter, we considered analysis of the effect of various functions of mass (specifically, mass, and mass<sup>2</sup>) on the survival of a hypothetical species of bird (the simulated data are in file `indcov1.inp`). Suppose you are interested in deriving both the estimated survival and the SE for this estimate, for an individual weighing some amount.

From Chapter 11, the linear function relating survival to mass and mass<sup>2</sup>, *on the logit scale*, is

$$\text{logit}(\phi) = 0.256733 + 1.1750545(\text{mass}_s) - 1.0555046(\text{mass}_s^2)$$

Note that for the two mass terms, there is a small subscript 's' - reflecting the fact that these are '*standardized*' masses. Recall that we standardized the covariates by subtracting the mean of the covariate, and dividing by the standard deviation (the use of standardized or non-standardized covariates is discussed at length in Chapter 11). Thus, for each individual in the sample, the estimated survival (on the logit scale) for that individual, given its mass, is given by

$$\text{logit}(\phi) = 0.256733 + 1.17505 \left( \frac{m - \bar{m}}{SD_m} \right) - 1.0555 \left( \frac{m^2 - \bar{m}^2}{SD_{m^2}} \right)$$

In this expression,  $m$  refers to mass and  $m^2$  refers to mass<sup>2</sup>. The output from **MARK** (preceding page) actually gives you the mean and standard deviations for both covariates: for mass, mean = 109.97, and SD = 24.79, while for mass<sup>2</sup>, the mean = 12707.46, and the SD = 5532.03. The 'value' column

shows the standardized values for mass and mass2 (0.803 and 0.752) for the first individual in the data file. Lets look at an example. Suppose the mass of the bird was 110 units. Thus mass = 110, mass2 =  $110^2 = 12100$ . Thus,

$$\text{logit}(\phi) = 0.2567 + 1.17505 \left( \frac{(110 - 109.97)}{24.79} \right) - 1.0555 \left( \frac{(12100 - 12707.46)}{5532.03} \right) = 0.374.$$

So, if  $\text{logit}(\phi) = 0.374$ , then how do we get the reconstituted values for survival? Recall that

$$\begin{aligned} \text{logit}(\theta) &= \log \left( \frac{\theta}{1 - \theta} \right) = \alpha + \beta x \\ \theta &= \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}} \end{aligned}$$

Thus, if  $\text{logit}(\phi) = 0.374$ , then the reconstituted estimate of  $\phi$ , transformed back from the logit scale is  $e^{0.374} / (1 + e^{0.374}) = 0.592$ . Thus, for an individual weighing 110 units, the expected annual survival rate is approximately 0.5925 (which is what **MARK** reports if you use the 'User specify covariate' option).

OK, but what about the variance (and corresponding SE) for this estimate? First, what is our 'transformation function' ( $Y$ )? Easy - it's the 'back-transform' of the linear equation on the logit scale. Given that

$$\begin{aligned} \text{logit}(\hat{\phi}) &= \beta_0 + \beta_1(\text{mass}_s) + \beta_2(\text{mass}_s^2) \\ &= 0.256733 + 1.1750545(\text{mass}_s) - 1.0555046(\text{mass}_s^2) \end{aligned}$$

then the back-transform function  $Y$  is

$$Y = \frac{e^{0.256733 + 1.1750545(\text{mass}_s) - 1.0555046(\text{mass}_s^2)}}{1 + e^{0.256733 + 1.1750545(\text{mass}_s) - 1.0555046(\text{mass}_s^2)}}$$

Second, since our transformation clearly involves multiple parameters ( $\beta_0, \beta_1, \beta_2$ ), the estimate of the variance is given by

$$\begin{aligned} \text{var}(\mathbf{Y}) &= \mathbf{D}\Sigma\mathbf{D}^T \\ &= \left( \frac{\partial(\hat{Y})}{\partial(\hat{\theta})} \right) \cdot \hat{\Sigma} \cdot \left( \frac{\partial(\hat{Y})}{\partial(\hat{\theta})} \right)^T \end{aligned}$$

Given our linear (transformation) equation (from above) then the vector of partial derivatives is (we've substituted  $m$  for mass and  $m2$  for mass2, and transposed it to make it easily fit on the page - it's pretty 'ugly'):

$$\begin{aligned} & \left[ \begin{array}{ccc} \left( \frac{\partial(\hat{Y})}{\partial\hat{\beta}_0} \right) & \left( \frac{\partial(\hat{Y})}{\partial\hat{\beta}_1} \right) & \left( \frac{\partial(\hat{Y})}{\partial\hat{\beta}_2} \right) \end{array} \right]^T \\ &= \left[ \begin{array}{ccc} \frac{e^{\beta_0 + \beta_1(m) + \beta_2(m2)}}{1 + e^{\beta_0 + \beta_1(m) + \beta_2(m2)}} - \frac{(e^{\beta_0 + \beta_1(m) + \beta_2(m2)})^2}{(1 + e^{\beta_0 + \beta_1(m) + \beta_2(m2)})^2} & & \\ \frac{m \times e^{\beta_0 + \beta_1(m) + \beta_2(m2)}}{1 + e^{\beta_0 + \beta_1(m) + \beta_2(m2)}} - \frac{m \times (e^{\beta_0 + \beta_1(m) + \beta_2(m2)})^2}{(1 + e^{\beta_0 + \beta_1(m) + \beta_2(m2)})^2} & & \\ \frac{m2 \times e^{\beta_0 + \beta_1(m) + \beta_2(m2)}}{1 + e^{\beta_0 + \beta_1(m) + \beta_2(m2)}} - \frac{m2 \times (e^{\beta_0 + \beta_1(m) + \beta_2(m2)})^2}{(1 + e^{\beta_0 + \beta_1(m) + \beta_2(m2)})^2} & & \end{array} \right] \end{aligned}$$

While this is, indeed, fairly ‘ugly’ looking, in fact the structure is quite straightforward - the only difference between the 3 elements of the vector is that the numerator of both terms (on either side of the minus sign) are multiplied by 1, m, and m2, respectively. Where do these scalar multipliers come from? Easy, they’re the partial derivatives of the linear model (we’ll call it  $f$ ) on the logit scale

$$f = \text{logit}(\hat{\phi}) = \beta_0 + \beta_1(m_s) + \beta_2(m_s^2)$$

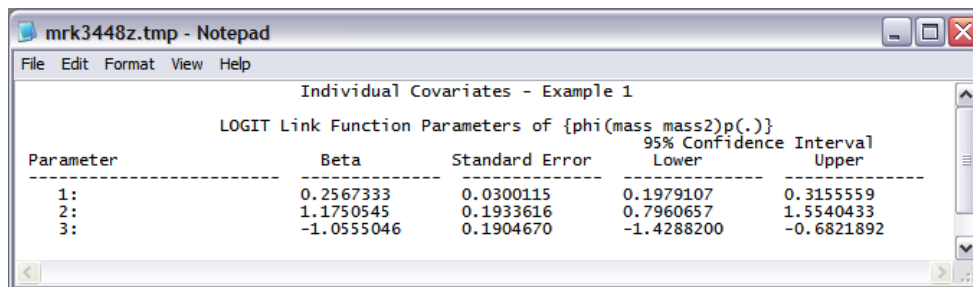
with respect to each of the parameters ( $\beta_i$ ) in turn. In other words,  $\partial f / \partial \beta_0 = 1$ ,  $\partial f / \partial \beta_1 = m$ , and  $\partial f / \partial \beta_2 = m2$ .

So, now that we have our vectors of partial derivatives of the transformation function with respect to each of the parameters, we can simplify things considerably by substituting in the the standardized values for m and m2, and the estimated parameter values ( $\hat{\beta}_0$ ,  $\hat{\beta}_1$ , and  $\hat{\beta}_2$ ).

For a mass of 110 g, the standardized values for mass and mass2 are

$$\text{mass}_s = \left( \frac{110 - 109.97}{24.79} \right) = 0.0012102 \quad \text{mass2}_s = \left( \frac{12100 - 12707.46}{5532.03} \right) = -0.109808$$

The estimates for  $\hat{\beta}_i$  we read directly from **MARK**:



Parameter	Beta	Standard Error	95% Confidence Interval	
			Lower	Upper
1:	0.2567333	0.0300115	0.1979107	0.3155559
2:	1.1750545	0.1933616	0.7960657	1.5540433
3:	-1.0555046	0.1904670	-1.4288200	-0.6821892

So, substituting in the estimates for  $\hat{\beta}_i$  and the standardized  $m$  and  $m_2$  values (from the previous page) into our vector of partial derivatives yields

$$\left[ \left( \frac{\partial(\hat{Y})}{\partial\hat{\beta}_0} \right) \quad \left( \frac{\partial(\hat{Y})}{\partial\hat{\beta}_1} \right) \quad \left( \frac{\partial(\hat{Y})}{\partial\hat{\beta}_2} \right) \right]^T = \begin{bmatrix} 0.24145 \\ 0.00029 \\ -0.02651 \end{bmatrix}$$

>From the **MARK** output (after exporting to a dBase file - and **not** to the Notepad - in order to get full precision), the full V-C matrix is

$$\begin{pmatrix} 0.0009006921 & -0.0004109710 & 0.0003662359 \\ -0.0004109710 & 0.0373887267 & -0.0364250288 \\ 0.0003662359 & -0.0364250288 & 0.0362776933 \end{pmatrix}$$

So,

$$\begin{aligned} \widehat{\text{var}}(\hat{Y}) &= \begin{bmatrix} 0.24145 & 0.00029 & -0.02651 \end{bmatrix} \\ &\times \begin{pmatrix} 0.0009006921 & -0.0004109710 & 0.0003662359 \\ -0.0004109710 & 0.0373887267 & -0.0364250288 \\ 0.0003662359 & -0.0364250288 & 0.0362776933 \end{pmatrix} \times \begin{bmatrix} 0.24145 \\ 0.00029 \\ -0.02651 \end{bmatrix} \\ &\approx 0.00007387 \end{aligned}$$

So, the estimated SE for  $\widehat{\text{var}}$  for the reconstituted value of survival for an individual weighing 110 g is  $\sqrt{0.00007387} = 0.00860$ , which is what is reported by **MARK** (again, within rounding error).

It is important to remember that the estimated variance will vary depending on the mass you use - the estimate of the variance for a 110 g individual (0.00007387) will differ from the estimated variance for a (say) 120 g Individual. For a 120 g individual, the standardized values of mass and mass<sup>2</sup> are 0.4045568999 and 0.3059519429, respectively. Based on these values, then

$$\begin{aligned} &\left[ \left( \frac{\partial(\hat{Y})}{\partial\hat{\beta}_0} \right) \quad \left( \frac{\partial(\hat{Y})}{\partial\hat{\beta}_1} \right) \quad \left( \frac{\partial(\hat{Y})}{\partial\hat{\beta}_2} \right) \right]^T \\ &= \begin{bmatrix} 0.23982 \\ 0.08871 \\ 0.07337 \end{bmatrix} \end{aligned}$$

Given the variance covariance-matrix for this model (shown above), then

$$\widehat{\text{var}}(\hat{Y}) = \mathbf{D}\Sigma\mathbf{D}^T \approx 0.000074214$$

Thus, the estimated SE for  $\widehat{\text{var}}$  for the reconstituted value of survival for an individual weighing 120 g is  $\sqrt{0.000074214} = 0.008615$ , which is what is reported by **MARK** (again, within rounding error).

Note that this value for the SE for a 120 g individual (0.008615) differs from the SE estimated for a 110 g individual (0.008600), albeit not by much (the small difference here is because this is a very large simulated data set based on a deterministic model - see Chapter 11 for details). Since each weight would have its own estimated survival, and associated estimated variance and SE, to generate a curve showing the reconstituted values and their SE, you'd need to iteratively calculate  $\mathbf{D}\Sigma\mathbf{D}^T$  over a range of weights. We'll leave it to you to figure out how to handle the programming if you want to do this on your own. For the less ambitious, **MARK** now has the capacity to do much of this for you - you can output the 95% CI 'data' over a range of individual covariate values to a spreadsheet (see section 11.5 in Chapter 11).

## B.5. Summary

In this appendix, we've briefly introduced a convenient, fairly straightforward method for deriving an estimate of the sampling variance for transformations of one or more variables. Such transformations are quite commonly encountered when using **MARK**, and having a method to derive estimates of the sampling variances is convenient. The most straightforward method - based on a first-order Taylor series expansion - is known generally as the 'Delta method'. However, as we saw, the first-order Taylor series approximation may not always be appropriate, especially if the transformation is highly non-linear, and if there is significant variation in the data. In such case, you may have to resort to higher-order approximations, or numerically intensive bootstrapping approaches.