Paper 164-2023

The Delta Method in Statistical Inference, with Applications in the SAS ® IML Procedure

Carter J Sevick, MS, Division of Health Care Policy and Research,
School of Medicine, University of Colorado Anschutz Medical Campus,
Aurora, Colorado, USA

Camille M Moore, PhD, Center for Genes, Environment and Health,
National Jewish Health, Denver, Colorado, USA

Samantha MaWhinney, ScD, Department of Biostatistics,
Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA

Abstract

Testing of substantive hypotheses may involve complex computations on estimated model parameters without clear, or immediate, standard errors. Possible examples include a difference in group proportions, from a multivariable logistic regression, or summary effects from a pattern mixture model. Inference can often be readily carried out using resampling methods, such as the bootstrap, but these strategies may be impractical in cases of large data, computationally intensive models or both. The Delta Method has a long and rich history in statistics and may be able to deliver valid and reliable inferences within acceptable time frames when resampling methods would not. In this work, the Delta Method is reviewed with attention to cases when it works well and when other strategies would be advisable. Included examples are graded from simple to complex, and detailed code using the IML procedure is provided.

The Delta Method

When attempting to make an inference on a parameter for which the distribution is unknown or difficult to work with, an estimate of the variance may be obtained using the Delta Method, provided the estimator can be written in terms of another with an asymptotically normal distribuition[1, 2].

Assume that:

$$\sqrt{n}(T_n - \theta) \xrightarrow{\mathcal{L}} N(0, \sigma^2)$$

Where $\xrightarrow{\mathcal{L}}$ indicates convergence in law, and T_n is a sequence of random variables with T_1 having mean θ and variance σ^2. let g be a function such that its first derivative evaluated at θ exists and is non-zero, then:

$$\sqrt{n}\big(g(T_n) - g(\theta)\big) \xrightarrow{\mathcal{L}} N(0, \sigma^2 g'(\theta)^2)$$

Essentially, $g(T_n)$ is distributed asymptotically normal with mean $g(\theta)$ and variance $\frac{\sigma^2 g'(\theta)^2}{n}$.

Example 1: Asymptotic Variance of the Logit

Let X_i, $i = 1, \ldots, n$, be independent and identically distributed (iid) as Bernoulli random variables with probability of success p. The mean of the X_i, $\hat{p}$, is the maximum likelihood estimator (MLE) for p, thus by the asymptotic efficiency of MLEs[2]:

$$\sqrt{n}(\hat{p} - p) \xrightarrow{\mathcal{L}} N\big(0, p(1-p)\big)$$

Now, consider the logit transformation: $\phi(p) = log\left(\frac{p}{1-p}\right)$. The first derivative is:

$$\frac{d\phi(p)}{dp} = \frac{d}{dp} log\left(\frac{p}{1-p}\right) = \frac{d}{dp}[log(p) - log(1-p)] = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}.$$

This derivative exists, and is non-zero, as long as p is not zero or one. It then follows from the Delta Method that

$$\sqrt{n}(\phi(\hat{p}) - \phi(p)) \xrightarrow{\mathcal{L}} N\left(0, \frac{1}{p(1-p)}\right)$$

How does this help the applied analyst? When we are trying to make an inference about a parameter from study data we estimate the distribution of T_n by the plug-in principle and use extension for the distribution of $g(T_n)$[4] and can construct significance tests and confidence intervals.

Example 2, continuation of example 1:

The Wald confidence interval for p is known to have poor coverage and may include impossible values when the sample size is not large and p is close to zero or one. The logit transform converges to a normal distribution more quickly than the identity, resulting in a more robust confidence interval[3]. In example 1 we used the Delta Method to derive the asymptotic distribution of $\phi(\hat{p}) = \phi\left(\frac{\sum_{i=1}^{n} X_i}{n}\right)$, and using $\hat{p}$ as a plug-in estimate for p, it follows that a $1-\alpha$ confidence interval for $\phi(p)$ is:

$$[L(\hat{p}), U(\hat{p})] = \left[\phi(\hat{p}) - Z_{1-\alpha/2}\left(\frac{1}{n\hat{p}(1-\hat{p})}\right)^{1/2}, \phi(\hat{p}) + Z_{1-\alpha/2}\left(\frac{1}{n\hat{p}(1-\hat{p})}\right)^{1/2}\right],$$

where $Z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. Then, by applying the inverse logit transformation $\left(\frac{1}{1+exp(-x)}\right)$ a $1-\alpha$ confidence interval for p is:

$$\left[\frac{1}{1+exp\left(-L(\hat{p})\right)}, \frac{1}{1+exp\left(-U(\hat{p})\right)}\right].$$

The Multivariate Delta Method

Univariate analysis can be useful, but fortunately the Delta Method has been extended to the multivariate case[1].

Assume that:

$$\sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \boldsymbol{\Sigma})$$

Where $\mathbf{T}_n$ is a sequence of random vectors of dimension m, with $\mathbf{T}_1$ having mean vector $\boldsymbol{\theta}$**,** also dimension m, and variance matrix $\boldsymbol{\Sigma}$ (m *by* m). let $\mathbf{G}$ be a real, m-vector, valued transformation, $\mathbf{G}$, with non-singular partial derivative (with respect to $\boldsymbol{\theta}$) matrix $\mathbf{B}$**,** then:

$$\sqrt{n}\left(\mathbf{G}(\mathbf{T}_n) - \mathbf{G}(\boldsymbol{\theta})\right) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}')$$

Essentially, $\mathbf{G}(\mathbf{T}_n)$ is distributed asymptotically multivariate normal with mean vector $\mathbf{G}(\boldsymbol{\theta})$ and variance $\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}'n^{-1}$.

Example 3, interpreting diagnostic tests with likelihood ratios:

SAS Usage Note 24170 presents an analysis of a screening test vs some gold standard result. In this context, the likelihood ratio of a positive result ($LR+$) is the ratio of the sensitivity to one minus the specificity, or the ratio of the probability of a positive test given disease to the probability of a positive test when disease free. The data from the study is:

Table 1. Evaluation Data for a Screening Test
Test, truth

+,+	+,-	-,+	-,-
11	4	2	6

The $LR+$ is calculated as:

$$\frac{11/(11+2)}{4/(4+6)} \approx 2.1$$

Assuming that only the sample size was fixed, results such as this can be considered as a sample of size n from a multinomial distribution with parameter vector $\mathbf{P} = [p_{++}, p_{+-}, p_{-+}, p_{--}]^T$, with the constraint that the sum of the elements equals one. Realized cell proportions will have a distribution that converges to a multivariate normal distribution with mean vector $\mathbf{P}$ and covariance matrix:

$$\mathbf{V}_P = (diag(\mathbf{P}) - \mathbf{P}\mathbf{P}^T)n^{-1} = \begin{pmatrix} p_{++}(1-p_{++}) & -p_{++}p_{+-} & -p_{++}p_{-+} & -p_{++}p_{--} \\ -p_{+-}p_{++} & p_{+-}(1-p_{+-}) & -p_{+-}p_{-+} & -p_{+-}p_{--} \\ -p_{-+}p_{++} & -p_{-+}p_{+-} & p_{-+}(1-p_{-+}) & -p_{-+}p_{--} \\ -p_{--}p_{++} & -p_{--}p_{+-} & -p_{--}p_{-+} & p_{--}(1-p_{--}) \end{pmatrix} n^{-1}$$

To derive the covariance matrix for the LR+, we will need to apply two transformations, one of which will need the Delta Method: once to create a vector that includes both the numerators and denominators of the sensitivity and 1-specificity and second to calculate the LR+. For the first phase, we define a matrix of constants ($\mathbf{C}$) such that $\mathbf{CP} = \mathbf{P}'$ where $\mathbf{P}'$ has the quantities we need. In this case :

$$\mathbf{C}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

Then $\mathbf{C}_1\mathbf{P}$ is:

$$\mathbf{P}_1 = \mathbf{C}_1\mathbf{P} = \begin{pmatrix} p_{++} \\ p_{++} + p_{-+} \\ p_{+-} \\ p_{+-} + p_{--} \end{pmatrix}$$

To calculate LR+, recall than logarithms can change multiplication to addition:

$$log(LR+) = log\left(\frac{p_{++}/(p_{++}+p_{-+})}{p_{+-}/(p_{+-}+p_{--})}\right) = log(p_{++}) - log(p_{++}+p_{-+}) - log(p_{+-}) + log(p_{+-}+p_{--}).$$

In matrix terms this is:

$$log(LR+) = \mathbf{C}_2 log(\mathbf{P}_1) = [1 \quad -1 \quad -1 \quad 1] log(\mathbf{P}_1).$$

Recalling that $\frac{\partial log(x)}{\partial x} = x^{-1}$, the matrix of partial derivatives, with respect to the elements of $\mathbf{P}$ is:

$$\frac{\partial \mathbf{C}_2 log(\mathbf{P}_1)}{\partial \mathbf{P}} = \mathbf{C_2 D}^* = \mathbf{C}_2 \begin{pmatrix} p_{++}^{-1} & 0 & 0 & 0 \\ (P_{++} + p_{+-})^{-1} & 0 & (P_{++} + p_{+-})^{-1} & 0 \\ 0 & 0 & p_{-+}^{-1} & 0 \\ 0 & (p_{+-} + p_{--})^{-1} & 0 & (p_{+-} + p_{--})^{-1} \end{pmatrix}$$

$$= \mathbf{C}_2 \begin{pmatrix} p_{++}^{-1} & 0 & 0 & 0 \\ 0 & (P_{++} + p_{+-})^{-1} & 0 & 0 \\ 0 & 0 & p_{-+}^{-1} & 0 \\ 0 & 0 & 0 & (p_{+-} + p_{--})^{-1} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} = \mathbf{C_2 D C}_1$$

Where $\mathbf{D}$ is a diagonal matrix of the elements of $\mathbf{P}_1$.

Finally, the asymptotic variance of log(LR+) is:

$$V_2 = \mathbf{C_2 D C}_1 \mathbf{V} (\mathbf{C_2 D C}_1)^T$$

After substituting estimates for the unknown parameters, a $1 - \alpha$ confidence interval for $LR+= exp\big(log(LR+)\big)$ is:

$$\big[exp\big(log\big(\widehat{LR+}\big) - Z_{1-\alpha/2}\hat{V}_2^{1/2}\big), exp\big(log\big(\widehat{LR+}\big) + Z_{1-\alpha/2}\hat{V}_2^{1/2}\big)\big]$$

Implementation in the IML Procedure

```
data scrnDat;
input Test Response Count;
datalines;
1 1 11
1 0 4
0 1 2
0 0 6
;

proc iml;

* read in study data to a vector ;
use scrnDat;
read all var {Count};
close scrnDat;
```

```
print Count;
```

Count
11
4
2
6

```
* compute estimated cell probabilities ;
N = sum(Count);
P = Count/N;

print P;
```

P
0.4782609
0.173913
0.0869565
0.2608696

```
* covariance matric of P ;
Vp = (diag(P) - P * P`)/N;

print Vp;
```

Vp			
0.010849	-0.003616	-0.001808	-0.005425
-0.003616	0.0062464	-0.000658	-0.001973
-0.001808	-0.000658	0.003452	-0.000986
-0.005425	-0.001973	-0.000986	0.0083833

```
* define matrix transformations ;
C1 = {1 0 0 0,
      1 0 1 0,
      0 1 0 0,
      0 1 0 1};

C2 = {1 -1 -1 1};

D = diag((C1 * P)##-1);

print D;
```

D			
2.0909091	0	0	0
0	1.7692308	0	0
0	0	5.75	0
0	0	0	2.3

```
* LR+ ;
logLRp = C2 * log(C1 * P);
LRp = exp(logLRp);


* the variance of log(LR+) ;
B = C2 * D * C1;
```

```
print B;
```

B			
0.3216783	-3.45	-1.769231	2.3

```
V = B * Vp * B`;


* 95% CI for LR+ ;
Z = quantile('normal', 0.975, 0, 1);
MOE = Z * sqrt(V);
LCL = exp(logLRp - MOE);
UCL = exp(logLRp + MOE);

print LRp[F = 8.2] LCL[F = 8.2] UCL[F = 8.2];
```

LRp	LCL	UCL
2.12	0.96	4.68

```
quit;
```

It is interesting to note that this confidence interval is calculated as 0.44 to 3.79 in SAS Usage Note 24170. The root of the issue is that the bounds are calculated from the estimated distribution of the ratio which converges to normality more slowly than the log-ratio. Reprogramming NLMIXED to estimate the log-LR+, and back-transforming, yields the same result that we present here.

Example 4: Average Risk Difference from a Logistic Regression

Sometimes an analysis may call for a summary statistic not produced by the model that fits the data well. For example, a logistic regression may be the best fitting model, but the investigator would like to interpret a relative risk or risk difference. In linear regression computing a prediction at covariate means is the same as the mean prediction across subjects, but this is not the case in a nonlinear model such as logistic regression[12]. One method of making the conversion is to compute predicted outcomes on a study dataset under different states and then contrast the average predictions[6,7,8]. This is sometimes referred to as the average predicted value (APV) method. A bootstrap or jackknife analysis can easily provide standard errors for inference, but may be computationally prohibitive with large data, complex models, or both. The Delta Method can be an efficient option, as long as the model fits the data well.

In this example we will use the low birth weight data[9] from Applied Logistic Regression, by Hosmer and Lemeshow. The set is composed of 189 deliveries and includes birth weight and various maternal factors. For this example we will estimate the odds of a low weight birth in mothers who smoke during pregnancy, vs those that do not, adjusted for maternal age. The primary statistic of interest will be the average risk difference between smokers and non.

```
data lowbwt;
  infile 'C:\Users\cjsev\Documents\wuss2023\lowbwt.txt' firstobs = 2 dlm='09'x;
  input ID LOW AGE LWT RACE SMOKE PTL HT UI FTV BWT;
run;


ods select OddsRatios;
proc logistic data  = lowbwt desc;
  model low = smoke age  ;
  store low_mod;
run;
```

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SMOKE	1.997	1.063	3.753
AGE	0.951	0.894	1.013

Based on this analysis, smoking during pregnancy may be associated with a 2 fold increase in the odds of a low birth weight delivery with a 95% confidence interval of 1.06 to 3.75.

To compute the average risk difference, let $\mathbf{X}_i$ be the covariate vector for subject i under a condition of interest (i.e. smoking during pregnancy) and $\mathbf{X}_i^*$ be the covariate vector of the same individual under an alternative condition (i.e. not smoking during pregnancy). The LOGISTIC procedure provides estimates via maximum likelihood (ML) and, assuming that the data are independent and identically distributed (iid), we have that the estimates of the model parameters, $\widehat{\mathbf{B}}$, are asymptotically multivariate normal with mean $\mathbf{B}$ and covariance matrix $\mathbf{V}$, by the asymptotic efficiency of maximum likelihood estimators[2]. Beginning with $\mathbf{B}$ known, the average risk difference between the two conditions can be calculated as:

$$\gamma = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{e^{\mathbf{X}_i\mathbf{B}}}{1+e^{\mathbf{X}_i\mathbf{B}}} - \frac{e^{\mathbf{X}_i^*\mathbf{B}}}{1+e^{\mathbf{X}_i^*\mathbf{B}}}\right)$$

The derivative of the expression is:

$$\boldsymbol{\gamma}' = \frac{\partial}{\partial\mathbf{B}}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{e^{\mathbf{X}_i\mathbf{B}}}{1+e^{\mathbf{X}_i\mathbf{B}}} - \frac{e^{\mathbf{X}_i^*\mathbf{B}}}{1+e^{\mathbf{X}_i^*\mathbf{B}}}\right)\right]$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial}{\partial\mathbf{B}}\frac{e^{\mathbf{X}_i\mathbf{B}}}{1+e^{\mathbf{X}_i\mathbf{B}}} - \frac{\partial}{\partial\mathbf{B}}\frac{e^{\mathbf{X}_i^*\mathbf{B}}}{1+e^{\mathbf{X}_i^*\mathbf{B}}}\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{e^{\mathbf{X}_i\mathbf{B}}}{(1+e^{\mathbf{X}_i\mathbf{B}})^2}\mathbf{X}_i - \frac{e^{\mathbf{X}_i^*\mathbf{B}}}{\left(1+e^{\mathbf{X}_i^*\mathbf{B}}\right)^2}\mathbf{X}_i^*\right)$$

For a model with p parameters this will be a $1xp$ vector. Then, by the Delta Method, the asymptotic variance is:

$$\boldsymbol{\gamma}'\mathbf{V}\boldsymbol{\gamma}'^{T}$$

Substituting estimates $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{V}}$ into the expressions, above, a $1-\alpha$ confidence interval for γ is:

$$\left[\hat{\gamma} - Z_{1-\alpha/2}\hat{\boldsymbol{\gamma}}\hat{\mathbf{V}}\hat{\boldsymbol{\gamma}}^T, \hat{\gamma} + Z_{1-\alpha/2}\hat{\boldsymbol{\gamma}}\hat{\mathbf{V}}\hat{\boldsymbol{\gamma}}^T\right]$$

The following program implements the procedure in IML:

```
* Extract model parameter and covariance estimates ;
ods select none;
ods output ParameterEstimates = parms
           cov = vcov
               ;
proc plm restore=low_mod;
 show parameters covariance;
run;
ods select all;

proc iml;

* Read covariance and paramter estimates into IML ;
** Covariance matrix ;
use vcov;
read all var {col1 col2 col3} into V;
close vcov;
```

```
print V;
```

V		
0.5735334	-0.047329	-0.023207
-0.047329	0.1035592	0.0000284
-0.023207	0.0000284	0.0010222

```
** Paramter estimates ;
use parms;
read all var {estimate} into B;
close parms;

print B;
```

B
0.0609051
0.6918486
-0.049779

```
* Covariate data ;
use lowbwt ;
read all var {age} into X ;
close lowbwt ;

* row count ;
N = nrow(X);

* Covariate matrix, all subjects smoking ;
X1 = j(N, 1, 1) || j(N, 1, 1) || X;

* print 5 rows of X1 ;
print (X1[1:5, ]);
```

1	1	28
1	1	29
1	1	34
1	1	25
1	1	25

```
* Covariate matrix, all subjects non-smoking ;
X0 = j(N, 1, 1) || j(N, 1, 0) || X;

* print 5 rows of X0 ;
print (X0[1:5, ]);
```

1	0	28
1	0	29
1	0	34
1	0	25
1	0	25

```
* Vector for log odds and scalar average predicted probabilities, smokers ;
XB1 = X1*B;
mu1 = mean(exp(XB1)/(1+exp(XB1)));

* Vector for log odds and scalar average predicted probabilities, non-smokers ;
XB0 = X0*B;
mu0 = mean(exp(XB0)/(1+exp(XB0)));
```

```
* Average risk difference ;
RD = mu1 - mu0;


* Derivative vector (Xm) ;
C1 = exp(XB1)/(1+exp(XB1))##2;
C0 = exp(XB0)/(1+exp(XB0))##2;
Xi = c1#x1-c0#x0;
Xm = mean(Xi);

print Xm;
```

Xm		
0.0495696	0.2366949	1.2050263

```
* variance of the risk difference, Delta Method ;
varm = Xm * V * Xm`;

* 95% CI for risk difference ;
Z = quantile('normal', 0.975, 0, 1);
MOE = Z * sqrt(varm);
LCL = RD - MOE;
UCL = RD + MOE;

print RD[F = 8.2] LCL[F = 8.2] UCL[F = 8.2];
```

RD	LCL	UCL
0.15	0.01	0.28

```
quit;
```

Maternal smoking was found to be associated with a 15 percentage point increase in the risk of a low birth weight delivery, on average. This estimate ranged from 1 to 28 points, with 95% confidence.

Example 5: Summarizing Average Effects from a Two-Part Model

Suppose we are studying cost of care for an insurance system during a month. It is unlikely that every client will have a medical expense during this time and the resulting data is likely to be zero heavy. One method of analysis would be to define a two-part model[10], one to estimate the probability that a cost is incurred and a second to estimate mean cost given a cost occurred. Interpretation of this model is difficult when estimates of effects on overall cost are desired, however the APV method can be used here as well. Using each part of the model a prediction of cost probability and mean cost is made, then the expected cost for an individual is just the product of the two predictions. Averaging across all study subjects we arrive at average cost predictions. As in example 4, covariate values can be manipulated to observe estimated effects on the population, which can be compared to alternative states. In this example it is assumed that the probability of a cost is modeled by a logistic regression, and the mean non-zero cost by a gamma GLM regression with a log link.

The data for this example is randomly generated by the following code:

```
data sdata ;

  call streaminit(12345);

  do id = 1 to 1000;

    ccc = rand('bernoulli', 0.5);
    prev12_lstd = rand('normal',  ccc-0.5, 1);

    p_cost = rand('bernoulli', 1/(1 + exp(-(log(0.25) + 0.55*ccc + 0.35*prev12_lstd))) );

    cost = 0;
    if p_cost = 1 then do;
      cost = rand('gamma', 1, exp(5 + 0.4*ccc + 0.45*prev12_lstd) /1);
    end;
```

```
    output;

  end;

  label ccc          = 'Indicator of a chronic condition (1/0)'
        prev12_lstd  = "Function of the subject's medical cost in the prior 12 months"
        p_cost       = 'Indicator of a cost in the current month'
        cost         = 'Cost in the current month';


run;
```

The probability and non-zero cost models are:

```
ods select OddsRatios;
ods output ParameterEstimates = p_parms
           covb = p_vcov;
proc logistic data = sdata ;
  model p_cost(event='1') = ccc prev12_lstd / covb;
run;
```

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
ccc	1.448	1.064	1.971
prev12_lstd	1.419	1.221	1.649

```
ods select ParameterEstimates;
ods output ParameterEstimates = c_parms
           covb = c_vcov;
proc genmod data = sdata ;
  where cost>0;
  model cost = ccc prev12_lstd / dist = gamma link = log covb ;
run;
```

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	4.9265	0.1052	4.7204	5.1326	2194.61	<.0001
ccc	1	0.5697	0.1467	0.2821	0.8573	15.07	0.0001
prev12_lstd	1	0.3182	0.0641	0.1926	0.4438	24.65	<.0001
Scale	1	0.8966	0.0694	0.7704	1.0434		

About half of the sample has a chronic condition and 26% have a non-zero cost in the month. Overall mean cost is about $56 (including zeros) and the goal is to study the effect of having a chronic condition on cost, adjusted by a function of the cost in the prior year.

Both the probability and non-zero cost models are estimated by ML, with iid data, thus the distribution of model parameter estimates are asymptotically multivariate normal[2]. Let $\overset{a}{\sim}$ denote asymptotically distributed.

Probability model:

$$\hat{\mathbf{B}}_p \overset{a}{\sim} N(\mathbf{B}_p, \mathbf{V}_p)$$

where $\hat{\mathbf{B}}_p$ and $\mathbf{B}_p$ are $1xr$ vectors and $\mathbf{V}_p$ is a rxr covariance matrix.

Cost model:

$$\widehat{\mathbf{B}}_c \overset{a}{\sim} N(\mathbf{B}_c, \mathbf{V}_c)$$

Where $\widehat{\mathbf{B}}_c$ and $\mathbf{B}_c$ are $1xq$ vectors and $\mathbf{V}_c$ is a qxq covariance matrix.

The average prediction for an individual, assuming $\mathbf{B}_p$ and $\mathbf{B}_c$ are known, can be computed as:

$$\phi(\mathbf{X}_{pi}, \mathbf{B}_p, \mathbf{X}_{ci}, \mathbf{B}_c) = \frac{e^{\mathbf{X}_{pi}\mathbf{B}_p}}{1 + e^{\mathbf{X}_{pi}\mathbf{B}_p}} e^{\mathbf{X}_{ci}\mathbf{B}_c}.$$

Where $\mathbf{X}_{pi}$ is a $1xr$ vector of covariate values for the probability model and $\mathbf{X}_{ci}$ is a $1xq$ vector of covariate values for the cost model.

The derivative, with respect to the model parameters is:

$$\phi'(\mathbf{X}_{pi}, \mathbf{B}_p, \mathbf{X}_{ci}, \mathbf{B}_c) = \left[\frac{e^{\mathbf{X}_{pi}\mathbf{B}_p}}{\left(1 + e^{\mathbf{X}_{pi}\mathbf{B}_p}\right)^2} e^{\mathbf{X}_{ci}\mathbf{B}_c}\mathbf{X}_{pi}, \quad \frac{e^{\mathbf{X}_{pi}\mathbf{B}_p}}{1 + e^{\mathbf{X}_{pi}\mathbf{B}_p}} e^{\mathbf{X}_{ci}\mathbf{B}_c}\mathbf{X}_{ci}\right]$$

Which is a $1x(r+q)$ vector. Letting a super-scripted '*', again, indicate a covariate vector adjusted to an alternative state, the average cost difference is:

$$\gamma = \frac{1}{N}\sum_{i=1}^{N}\left(\phi(\mathbf{X}_{pi}, \mathbf{B}_p, \mathbf{X}_{ci}, \mathbf{B}_c) - \phi(\mathbf{X}^*_{pi}, \mathbf{B}_p, \mathbf{X}^*_{ci}, \mathbf{B}_c)\right),$$

and the derivative is:

$$\boldsymbol{\gamma}' = \frac{1}{N}\sum_{i=1}^{N}\left(\phi'(\mathbf{X}_{pi}, \mathbf{B}_p, \mathbf{X}_{ci}, \mathbf{B}_c) - \phi'(\mathbf{X}^*_{pi}, \mathbf{B}_p, \mathbf{X}^*_{ci}, \mathbf{B}_c)\right),$$

a $1x(r+q)$ vector.

Now, let the covariance matrices be $\mathbf{V}_p$ and $\mathbf{V}_c$ for the probability and cost models. The two-part model assumes that the two parts are independent, thus the combined covariance matrix is block diagonal:

$$\mathbf{V}_{pc} = \begin{bmatrix}\mathbf{V}_c & 0\\ 0 & \mathbf{V}_c\end{bmatrix}.$$

Then the variance of the cost difference is:

$$\boldsymbol{\gamma}'\mathbf{V}_{pc}\boldsymbol{\gamma}'^{T}.$$

Substituting estimates $\widehat{\mathbf{B}}_p$, $\widehat{\mathbf{B}}_c$ and $\widehat{\mathbf{V}}_{pc}$ into the expressions, above, a $1-\alpha$ confidence interval for γ is:

$$\left[\hat{\gamma} - Z_{1-\alpha/2}\hat{\boldsymbol{\gamma}}\widehat{\mathbf{V}}_{pc}\hat{\boldsymbol{\gamma}}^T, \hat{\gamma} + Z_{1-\alpha/2}\hat{\boldsymbol{\gamma}}\widehat{\mathbf{V}}_{pc}\hat{\boldsymbol{\gamma}}^T\right]$$

The full process is implemented in the following code:

```
proc iml;

* read in estimated parameter vectors;
use p_parms;
read all var {estimate} into Bp;
close p_parms;

use c_parms(where = (parameter NE 'Scale'));
read all var {estimate} into Bc;
close c_parms;

* read in co-variance matricies;
use p_vcov;
read all var {Intercept ccc prev12_lstd} into Vp;
close p_vcov;
```

```
use c_vcov(where = (rowname NE 'Scale'));
read all var {prm1 prm2 prm3} into Vc;
close c_vcov;

* get covariate matrix ;
use sdata ;
read all var {ccc prev12_lstd} into X_ ;
close sdata ;

* sample size ;
N = nrow(X_);

* X with intercept ;
X = j(N, 1, 1) || X_;

* Sample as all subjects having a chronic condition "exposed"  ;
X1 = j(N, 1, 1) || j(N, 1, 1) || X[, 3];

* Sample as no subjects having a chronic condition "unexposed";
X0 = j(N, 1, 1) || j(N, 1, 0) || X[, 3];

* "exposed" overall mean cost ;
X1Bp = X1*Bp;
P1   = exp(X1Bp)/(1+exp(X1Bp));
X1Bc = X1*Bc;
C1   = exp(X1*Bc);
M1 = mean(P1 # C1);

* "unexposed" overall mean cost ;
X0Bp = X0*Bp;
P0   = exp(X0Bp)/(1+exp(X0Bp));
X0Bc = X0*Bc;
C0   = exp(X0*Bc);
M0 = mean(P0 # C0);

* Average predicted cost differnce ;
M_diff = M1 - M0;

* derivative of the difference ;
D_diff = mean( ((P1/(1+exp(X1Bp)) # C1 # X1) || (P1 # C1 # X1))  -  ((P0/(1+exp(X0Bp)) #
C0 # X0) || (P0 # C0 # X0)));

print D_diff;
```

D_diff					
27.575251	53.329345	11.312551	45.00392	79.722948	24.964025

```
* full model covariance matrix ;
Vpc = block(Vp, Vc);

print Vpc;
```

Vpc					
0.0125421	-0.013043	0.0010624	0	0	0
-0.013043	0.0247383	-0.003823	0	0	0
0.0010624	-0.003823	0.0058576	0	0	0
0	0	0	0.0110589	-0.011703	0.0009733
0	0	0	-0.011703	0.0215323	-0.003692
0	0	0	0.0009733	-0.003692	0.0041084

```
* variance of the difference ;
Vd = D_diff * Vpc * D_diff`;


* 95% CI for Diff ;
Z = quantile('normal', 0.975, 0, 1);
MOE = Z * sqrt(Vd);
LCL = M_diff - MOE;
UCL = M_diff + MOE;

print M_diff[F = 8.2] LCL[F = 8.2] UCL[F = 8.2];
```

M_diff	LCL	UCL
45.00	25.05	64.96

```
Quit;
```

In this population having a chronic condition is associated with an estimated higher mean cost of $45, in a given month, which ranges from $25 to $65, with 95% confidence.

Cautions when using the Delta Method

In order to get standard error estimates comparable to the bootstrap the model must fit the data reasonably well. If the variances estimated by the model are incorrect then computations via the Delta Method will be similarly off.

Care should be taken that sample sizes are large enough that the central limit theorem (CLT) can be assumed to be in effect. For example, while differences tend to converge to normality quickly ratios tend to go more slowly, so if we were calculating the relative risk from a logistic model it is best to compute estimates and intervals for the log ratio and then exponentiate, as this transformation tends to normality quicker in this context (see example 3). A similar issue arises in cases of sparsity in high dimensional models[11] and the CLT may not hold where the number of estimated parameters is high relative to the sample size.

Finally, if the derivative of the transformation equals zero or does not exist when evaluated at the mean parameter then the Delta Method will fail. An extension to second order asymptotics exists which may fix the problem assuming that the second derivative of the transformation evaluated at the mean parameter exists and is non-zero[2]. This is of particular concern when the transformation is nonmonotonic.

Conclusions

Computing power continues to become more affordable and available, but the complexity of our questions and the size of our datasets continue to grow as well. The Delta Method is a tool that was developed to give reasonable answers to complex questions in an environment of computing scarcity. While resampling methods will continue to hold a prime place in the statistician's toolbox for difficult inferences, the Delta Method may still be of use when computation burdens are impractical.

References

1. Lehmann E. L., Casella G. (1998) "Theory of Point Estimation, 2nd Edition" Springer New York, NY
2. Casella G., Berger R. L. (2002) "Statistical Inference, 2nd Edition" Duxbury
3. Brown, L. D., Cai, T. T., & DasGupta, A. (2001). "Interval Estimation for a Binomial Proportion." Statistical Science, 16(2), 101–117. http://www.jstor.org/stable/2676784GSK
4. Bickel P. J., Doksum K. A., (2015) "Mathematical Statistics: Basic Ideas and Selected Topics, Volume 1" Chapman and Hall / CRC
5. Agresti A. (1990) "Categorical Data Analysis" John Wiley and Sons, inc.

6. Kleinman LC, Norton EC. (2009) “What's the Risk? A simple approach for estimating adjusted risk measures from nonlinear models including logistic regression.” Health Serv Res. Feb;44(1):288-302.
7. King G., Tomz M., Wittenberg J. (2000) “Making the Most of Statistical Analyses: Improving Interpretation and Presentation” American Journal of Political Science, 44, Pp. 341-355
8. Greenland S. (2004) “Model Based Estimation of Relative Risk and Other Epidemiologic Measures in Studies of Common Outcomes and in Case-Control Studies” American Journal of Epideiology 160: 4, Pp 301-305
9. Hosmer D. W., Lemeshow S. (2000) “Applied Logistic Regression, 2nd Edition” John Wiley and Sons, Inc.
10. Mihaylova B., Briggs A., O’Hagan A., Thompson S. G. (2011) “Review of Statistical Methods for Analysing Healthcare Resources and Costs” Health Economics. 20: Pp 897-916
11. Fitzmaurice G. M., Laird N. M. (2000) Generalized Linear Mixture Models for Handling Nonignorable Dropouts in Longitudinal Studies” Biostatistics, 1, 2, Pp. 141-156
12. Muller C. J., MacLehose R. F. (2014) “Estimating predicted probabilities from logistic regression: different methods correspond to different target populations” International Journal of Epidemiology, 43(3), 962–970.

Contact Information

Carter j Sevick

Carter.Sevick@cuanschutz.edu