Analysing error propagation using the delta method

June 6, 2011

Suppose that X and Y are random variables (e.g. estimators) and suppose that $Z = f(X,Y)$. Let $\mu_X = EX$ and $\mu_Y = EY$. If we approximate f by a first order taylor series in the two variables X and Y around the point $(x,y) = (\mu_X, \mu_Y)$ we get

$$Z = f(X,Y) \approx f(\mu_X, \mu_Y) + \left(\frac{\partial f}{\partial x}\right)(X - \mu_X) + \left(\frac{\partial f}{\partial y}\right)(Y - \mu_Y). \tag{1}$$

Subtracting $\mu_Z = f(\mu_X, \mu_Y)$ and squaring both sides of this equation yields

$$(Z - \mu_Z)^2 \approx (\left(\frac{\partial f}{\partial x}\right)(X - \mu_X))^2 + (\left(\frac{\partial f}{\partial y}\right)(Y - \mu_Y))^2 + 2\left(\frac{\partial f}{\partial x}\right)(X - \mu_X)\left(\frac{\partial f}{\partial y}\right)(Y - \mu_Y). \tag{2}$$

Taking expectations, we find that variance of Z is

$$\operatorname{Var} Z \approx \left(\frac{\partial f}{\partial x}\right)^2 \operatorname{Var} X + \left(\frac{\partial f}{\partial y}\right)^2 \operatorname{Var} Y + \left(\frac{\partial f}{\partial x}\right)\left(\frac{\partial f}{\partial y}\right) 2\operatorname{Cov}(X,Y). \tag{3}$$

This formula can be generalized to functions of any number of variables. Note that all partial derivates are taken at the point $(x,y) = (\mu_X, \mu_Y)$.

This approximation is typically used in situation where we are interested in some parameter being a function of other parameters for which the standard errors are known. Consider the logistic regression model on p. 240 in Dalgaard. Here a model for the probability that girls between 8 and 20 years have had their first menstrual cycle is analysed using the model

$$\operatorname{logit} p = \beta_0 + \beta_1 x, \tag{4}$$

where x is age. Suppose we want an estimate of the mean age x_0 at which the first menstrual cycle occurs. At this age, the probability that the first menstrual cycle already has occured $p = 1/2$, $\operatorname{logit} p = \ln(1/2/(1-1/2)) = 0$ and so

$$\beta_0 + \beta_1 x_0 = 0. \tag{5}$$

Solving for x_0, we see that x_0 is a parameter being a function of the parameters β_0 and β_1, namely

$$x_0 = f(\beta_0, \beta_1) = -\frac{\beta_0}{\beta_1}. \tag{6}$$

From the summary of the fitted logistic regression

```
> menmod <- glm(menarche ~ age,binomial)
> summary(menmod)

Call:
glm(formula = menarche ~ age, family = binomial)
```

```
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.32759  -0.18998   0.01253   0.12132   2.45922

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -20.0132     2.0284  -9.867   <2e-16 ***
age           1.5173     0.1544   9.829   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 719.39  on 518  degrees of freedom
Residual deviance: 200.66  on 517  degrees of freedom
AIC: 204.66

Number of Fisher Scoring iterations: 7
```

we have ML estimates $\hat{\beta}_0$ and $\hat{\beta}_1$. Functional invariance of ML estimates implies that the ML estimate of x_0 is

$$\hat{x}_0 = f(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\hat{\beta}_0}{\hat{\beta}_1}. \tag{7}$$

The estimate of x_0 is thus $\hat{x}_0 = -(-20.01)/1.51 = 13.25$ years.

The approximate variance (and the standard error) of this estimator can computed using (3). Keep in mind that all partial derivatives should be evaluated in the respective mean values equal to the above point estimates. These derivatives are thus

$$\begin{aligned}
\left.\frac{\partial f}{\partial \beta_0}\right|_{\substack{\beta_0=\hat{\beta}_0\\ \beta_1=\hat{\beta}_1}} &= -\left.\frac{1}{\beta_1}\right|_{\substack{\beta_0=\hat{\beta}_0\\ \beta_1=\hat{\beta}_1}} = -\frac{1}{\hat{\beta}_1} = -0.658\\
\left.\frac{\partial f}{\partial \beta_1}\right|_{\substack{\beta_0=\hat{\beta}_0\\ \beta_1=\hat{\beta}_1}} &= \left.\frac{\beta_0}{\beta_1^2}\right|_{\substack{\beta_0=\hat{\beta}_0\\ \beta_1=\hat{\beta}_1}} = \frac{\hat{\beta}_0}{\hat{\beta}_1^2} = -8.65
\end{aligned} \tag{8}$$

The variances of $\hat{\beta}_0$ and $\hat{\beta}_1$ are equal to the square of the standard errors in the summary. The estimated covariance between the estimators can be obtained using the function `vcov` on the fitted model object.

```
> vcov(menmod)
            (Intercept)         age
(Intercept)   4.1142802 -0.31188805
age          -0.3118881  0.02383202
```

This matrix contain the covariances between all the regression coefficients and the intercept. The estimated covariance betwen $\hat{\beta}_0$ and $\hat{\beta}_1$ is thus $\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) = -0.312$. The elements along the diagonal are the variances.

The approximate variance then become

$$\mathrm{Var}\,\hat{x}_0 \approx \left(\frac{\partial f}{\partial \beta_0}\right)^2 \mathrm{Var}\,\hat{\beta}_0 + \left(\frac{\partial f}{\partial \beta_1}\right)^2 \mathrm{Var}\,\hat{\beta}_1 + 2\left(\frac{\partial f}{\partial \beta_0}\right)\left(\frac{\partial f}{\partial \beta_1}\right)\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1). \tag{9}$$

If we attempt to compute this in R using three significant digits for all variances and covariances we get

```
> (-.658)^2*2.02^2 + (-8.65)^2*.154^2 + 2*(-.658)*(-8.65)*(-0.318)
[1] -0.0878712
```

that is, a negative variance. Surely this is wrong. This arise because the true covariance matrix of $(\hat{\beta}_0, \hat{\beta}_1)$ happen to be close to so called negative definite as seen by the high correlation of -0.996 between $\hat{\beta}_0$ and $\hat{\beta}_1$,

```
> cov2cor(vcov(menmod))
            (Intercept)        age
(Intercept)   1.0000000 -0.9960277
age          -0.9960277  1.0000000
```

which makes the numerical computations based on it potentially unstable if too few digits are used. The way around the problem is to store the the variance-covariance matrix in a new variable and do the computations with full double precision accuracy

```
> v <- vcov(menmod)
> v[1,1]*.658^2 + v[2,2]*8.65^2 + 2*v[1,2]*-.658*-8.65
[1] 0.01415985
> sqrt(v[1,1]*.658^2 + v[2,2]*8.65^2 + 2*v[1,2]*-.658*-8.65)
[1] 0.1189952
```

This gives a positive variance and $\mathrm{SE}(\hat{x}_0) = 0.119$ years. Note also that ignoring the last term involving the large negative covariance would lead to a huge error.