Determining the Confidence Interval for the Coefficient of Variation in a DBC Cluster

Maarten Buiter, Mark de Groot, Evelyn Lai

November 1, 2007

Abstract

1 Introduction

In order to limit the ever rising costs of healthcare in the Netherlands, the government has introduced the Diagnose-Behandel-Combinatie (DBC) system. This unit of entity in this system is the DBC which is a combination of a diagnose (D), and a number of treatments (Behandeling (B)) that logically follow the diagnose. The system was introduced to increase competition between hospitals as it makes clear how much each hospital charges for a particular DBC. The cheapest hospital will make profit, the most expensive one looses money.

However, the number of DBCs has risen dramatically. This led to a simplication of the system where comparable DBCs were put into clusters, which became the new unit of entity. A cluster of DBCs is considered homogeneous when the *coefficient of variation* (CV) of all DBCs in a cluster is less than 0.6.

We will use the delta-method and bootstrapping to determine the 95% confidence interval of the CV.

1.1 Distribution and likelihood

The costs of the DBCs in a cluster can be considered as taken from a normal distribution with mean μ and standard deviation σ.

The Density Function of the normal distribution is

$$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \tag{1}$$

The logarithm of the likelihood function, the loglikelyhood function l, is characterized by

$$l = \sum_{k=1}^{n} -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(x-\mu)}{2\sigma^2} \tag{2}$$

1.2 The Asymptotic Covariance Matrix

The two-by-two *asymptotic covariance matrix* contains the derivatives of the loglikelihood function with respect to μ and σ. The derivatives of the loglikelihood funtion with respect to (μ, μ), (μ, σ), and (σ, σ) are, respectively:

$$\frac{\partial^2 l}{\partial \mu^2} = \sum_{k=1}^{n} -\frac{1}{\sigma^2} = -\frac{n}{\sigma^2} \tag{3}$$

$$\frac{\partial^2 l}{\partial \mu \partial \sigma} = \sum_{k=1}^{n} \frac{2(\mu - x)}{\sigma^3} = \frac{n(n - 2\mu + 1)}{\sigma^3} \tag{4}$$

$$\frac{\partial^2 l}{\partial \sigma^2} = \sum_{k=1}^{n} -\frac{3x^2 - 6\mu x + 3\mu^2 - \sigma^2}{\sigma^4} = \frac{n(2n^2 + 3n(1 - 2\mu) + 6\mu^2 - 6\mu - 2\sigma^2 + 1}{\sigma^4} \tag{5}$$

When these derivatives are put into the asymptotic covariance matrix H, the matrix looks like

$$H = \begin{pmatrix} \frac{\partial^2 l}{\partial \mu^2} & \frac{\partial^2 l}{\partial \mu \partial \sigma} \\ \frac{\partial^2 l}{\partial \mu \partial \sigma} & \frac{\partial^2 l}{\partial \sigma^2} \end{pmatrix}$$

2 Estimating μ and σ using the bootstrap method

We have used the bootstrap method (N=1000) to estimate the 95% confidence interval for the CV of the given sample data (a group of 25 DBCs with fictive costs). The commands used to create this bootstrap are given in Appendix 2.

The estimated 95% confidence interval for the sample data is (0.55 - 0.95). The real CV for the sample data (0.77) lies within this confidence interval. We can therefore consider this particular cluster of DBCs homogeneous.

3 Conclusion

Our conclusion is...

Appendix 1 - Sample Data

DBC Number	Costs
1	38
2	79
3	94
4	115
5	263
6	402
7	535
8	548
9	749
10	754
11	775
12	801
13	802
14	844
15	849
16	1007
17	1088
18	1115
19	1346
20	1359
21	1883
22	1966
23	2164
24	2551
25	2846

The mean and standard deviation are $\overline{x} = 998.92$ and $\sigma = 767.84$

Appendix 2 - Bootstrap R commands

Bootstrap function (bootstrap) downloaded from CRAN repository (USA CA1)

```
> x = c(38, 79, 94, 115, 263, 402, 535, 548, 749, 754, 775, 801, 802,
844, 849, 1007, 1088, 1115, 1346, 1359, 1883, 1966, 2164, 2551, 2846)
> N = 1000
> theta = function(x) { sd(x) / mean(x) }
> perc0025 = function(x) { quantile(x, 0.025) }
> perc975 = function(x) { quantile(x, 0.975) }
> bootstrap(x, N, theta, func=perc0025)
.
.
.
[991] 0.5013028 0.6231071 0.8372442 0.7983070 0.6777156 0.9426073
0.6068895 0.8523207 0.6686826 0.8158544

$func.thetastar
     2.5%
0.5531579

> bootstrap(x, N, theta, func=perc975)
.
.
.
[991] 0.7328652 0.8339498 0.7149956 0.6144862 0.6525930 0.8006111
0.7468614 0.6304814 0.8375282 0.8572355

$func.thetastar
    97.5%
0.9548017
>
```

Appendix 3 - Derivatives

In this appendix are (some) steps in the calculations of the derivatives.

3.1 The first order derivative with respect to (μ)

The first part of the derivative of the loglikelihood funtion with respect to (μ) is equal to zero. For the second part applies the quotient rule.

$$\frac{\partial l}{\partial \mu} = \sum_{i=1}^{n} 0 \times \frac{(x_i - \mu)^2}{\sigma^2} + -\frac{1}{2}\frac{\sigma^2(-2x_i + 2\mu) - 0 \times (x_i^2 - 2x_i\mu + \mu^2)}{(\sigma^2)^2} \tag{6}$$

Eliminating non necessary parts of the derivative.

$$\frac{\partial l}{\partial \mu} = \sum_{i=1}^{n} -\frac{1}{2}\frac{\sigma^2(-2x_i + 2\mu)}{\sigma^4} \tag{7}$$

Dividing by σ^2.

$$\frac{\partial l}{\partial \mu} = \sum_{i=1}^{n} -\frac{1}{2}\frac{-2x_i + 2\mu}{\sigma^2} \tag{8}$$

Multiplying by $-\frac{1}{2}$.

$$\frac{\partial l}{\partial \mu} = \sum_{i=1}^{n} \frac{x_i - \mu}{\sigma^2} \tag{9}$$

3.2 The second order derivative with respect to (μ,μ)

The quotient rule applies for the calculation of the second derivative.

$$\frac{\partial^2 l}{\partial \mu^2} = \sum_{i=1}^{n} \frac{\sigma^2 \times -1 - 0 \times \sigma^2}{(\sigma^2)^2} \tag{10}$$

The sum of the whole is the sum of its parts and n times $-\sigma$ is equal to $-n\sigma$.

$$\frac{\partial^2 l}{\partial \mu^2} = \frac{-n\sigma^2 - 0}{\sigma^4} = \frac{-n}{\sigma^2} \tag{11}$$

3.3 The first order derivative with respect to (σ)

For the first part of the derivative of the loglikelihood function with respect to (σ) applies the product rule (Leibniz's Law), for the logaritm of the first part applies the chain rule and for the second part applies the quotient rule.

$$\frac{\partial l}{\partial \sigma} = \sum_{i=1}^{n} -0 \times \log(2\pi\sigma^2) + -\frac{1}{2}\frac{1}{2\pi\sigma^2}4\pi\sigma - 0 \times \frac{(x_i - \mu)^2}{\sigma^2} + -\frac{1}{2}\frac{\sigma^2 \times 0 - 2\sigma(x_i - \mu)^2}{\sigma^4} \tag{12}$$

Elimination non necessary parts of the derivative.

$$\frac{\partial l}{\partial \sigma} = \sum_{i=1}^{n} -\frac{1}{\sigma} + \frac{(x_i - \mu)^2}{\sigma^3} \tag{13}$$

3.4 The second order derivative with respect to (σ,σ)

The quotient rule applies for both parts of the first derivative.

$$\frac{\partial^2 l}{\partial \sigma^2} = \sum_{i=1}^{n} -\frac{\sigma \times 0 - 1 \times 1}{\sigma^2} + \frac{\sigma^3 \times 0 - 3\sigma^2 (x_i - \mu)^2}{\sigma^6} \tag{14}$$

Followed by eliminating the non necessary parts.

$$\frac{\partial^2 l}{\partial \sigma^2} = \sum_{i=1}^{n} \frac{1}{\sigma^2} + \frac{-3(x_i - \mu)^2}{\sigma^4} \tag{15}$$

This derivative can be simplified by applying the knowledge of σ^2.

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n} \tag{16}$$

$$n\sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 \tag{17}$$

Simplifying the derivative by using the knowledge of σ^2.

$$\frac{\partial^2 l}{\partial \sigma^2} = \frac{n}{\sigma^2} + \frac{-3n\sigma^2}{\sigma^4} \tag{18}$$

$$\frac{\partial^2 l}{\partial \sigma^2} = \frac{n}{\sigma^2} + \frac{-3n}{\sigma^2} = -\frac{2n}{\sigma^2} \tag{19}$$