An alternative expression for the constant $c_4[n]$ with desirable properties

Muñoz, J.F., Moya-Fernández, P.J., Álvarez, E., Blanco-Encomienda, F.J.[1]

Faculty of Economics and Business, Campus Cartuja s/n. 18071 Granada, Spain

Abstract

The constant $c_4[n]$ is commonly used in the construction of control charts and the estimation of process capability indices, where n denotes the sample size. Assuming the Normal distribution the unbiased estimator of the population standard deviation is obtained by dividing the sample standard deviation by the constant $c_4[n]$. An alternative expression for $c_4[n]$ is proposed, and the mathematical induction technique is used to prove its validity. Some desirable properties are described. First, the suggested expression provides the exact value of $c_4[n]$. Second, it is not a recursive formula in the sense it does not depend on the previous sample size. Finally, the value of $c_4[n]$ can be directly computed for large sample sizes. Such properties suggest that the proposed expression may be a convenient solution in computer programming, and it has direct applications in statistical quality control.

Keywords: bias, standard deviation, gamma function, control chart, process capability index

[1]jfmunoz@ugr.es

1. Introduction

The population standard deviation (σ) of a given variable X is a popular statistic in many disciplines. For instance, the parameter σ has a special relevance in statistical quality control, since the variability of production processes is traditionally associated with the standard deviation of the quality characteristic (see Mitra [1], Montgomery [2]). Accordingly, σ is used in different statistical techniques such as control charts (Chen [3], Chou et al. [4], Huang et al. [5], Ajadi and Riaz [6], Diko et al. [7]), process capability indexes (Chen and Chou [8], Liao [9], Polansky and Maple [10], Keshteli et al. [11]), acceptance sampling (Kasprikova and Klufa [12], Robertson et al.[13]), etc. The parameter σ is usually unknown in practice, and an estimator with desirable properties is required in this situation (see Mahmoud et al. [14], Muñoz-Rosas et al. [15], Chen et al. [16], Parry et al. [17]). Let $x_1, \ldots, x_n$ denote the values of X for a random sample with size n. The usual estimator of σ is the sample standard deviation $\widehat{\sigma} = (\widehat{\sigma}^2)^{1/2}$, where

$$\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \tag{1}$$

is the sample variance, and

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

is the sample mean. The term $n-1$ into equation (1) is the Bessel's correction, which is used to achieve an unbiased estimator, i.e, $E[\widehat{\sigma}^2] = \sigma^2$. Although $\widehat{\sigma}^2$ is an unbiased estimator of σ^2, the Jensen's inequality (Kuczma [18]) can be used to show that $\widehat{\sigma}$ is a biased estimator of σ. Assuming a Normal distribution, the unbiased estimator of σ (see Bolch [19], Cryer and Ryan

[20], Gurland and Tripathi [21]) is given by

$$\widehat{\sigma}_{c4} = \frac{\widehat{\sigma}}{c_4[n]},$$

where the constant $c_4[n]$ is defined as

$$c_4[n] = \frac{2^{1/2}}{(n-1)^{1/2}} \frac{\Gamma[n/2]}{\Gamma[(n-1)/2]}, \tag{2}$$

and

$$\Gamma[a] = \int_0^\infty x^{a-1} e^{-x} dx$$

is the Gamma function. An alternative expression for $c_4[n]$ that also provides its exact value is given by

$$c_4[n] = \begin{cases} \dfrac{c_4[2]}{(n-1)^{1/2}} \dfrac{2^{n-2}(n/2-1)!^2}{(n-2)!} & \text{if } n \text{ is even} \\ \dfrac{1}{(n-1)^{1/2} c_4[2]} \dfrac{(n-2)!}{2^{n-3}((n-3)/2)!^2} & \text{if } n \text{ is odd} \end{cases} \tag{3}$$

where $c_4[2] = (2/\pi)^{1/2}$. Note that $c_4[n]$ is a known constant in social sciences, and it has an especial relevance in statistical quality control. Many references tabulate the value of $c_4[n]$ for different values of n (see Montgomery [2]). Various statistical software can be used to compute $c_4[n]$, but its value may not be available in the case of large samples. This is due to the fact that the equation (2) depends on the quotient of two Gamma functions, and $\Gamma[a]$ tends to infinite as a increases. Similarly, the value of (3) can not be directly obtained for large sample sizes, since the factorial function also tends to infinite as its argument increases. For instance, both Microsoft Excel and the statistical software R (R Core Team [22]) give a numerical solution of $\Gamma[a]$ up to the value $a = 171.6144$, which implies that $c_4[n]$ based upon (2) can not be computed when $n > 343$. In the case of using (3), both Microsoft

Excel and R give a numerical solution of $c_4[n]$ up to $n = 172$. A possible solution is to use a recursive expression, i.e., the value of $c_4[n]$ may depend on $c_4[n-1]$ or previous values of n. A recursive expression is given by

$$c_4[n] = \frac{(n-2)^{1/2}}{(n-1)^{1/2} \quad c_4[n-1]} \tag{4}$$

However, recursive expressions are generally more time consuming in practice, for example, when programming, as can be seen in Section 3. Alternatively, approximations may be used in the case of large sample sizes. Some examples are (see Bolch [19], Gurland and Tripathi [21], Mitra [1], Montgomery [2], Huberts et al.[23]):

$$c_4[n] \cong 1 - \frac{1}{4n} - \frac{7}{32n^2} - \frac{19}{128n^3}$$

and

$$c_4[n] \cong \frac{4(n-1)}{4n-3}.$$

We present an analytical expression for $c_4[n]$ (see equation (7)) that provides its exact value, and can be directly computed for large sample sizes, since it does not depend on Gamma or factorial functions. Microsoft Excel and/or R can be easily used to compute the suggested expression under large sample sizes. Such programming details are available from the authors. The suggested expression may be a convenient solution when computer programming, since it is calculable for large sample sizes and less time consuming than recursive expressions.

2. A new expression for the constant $c_4[n]$

First, alternative expressions for $c_4[n]$ based on both even and odd values of n are derived. Second, we suggest an expression for $c_4[n]$ that can be used for any value of n. Finally, the mathematical induction technique is used to show the validity of the suggested expression.

The suggested expression for $c_4[n]$ is obtained by evaluating the equation (4) at the first values of n. For instance, if we evaluate the expression (4) up to $n = 8$ we obtain (SEE APPENDIX II FOR MORE DETAIL):

$$c_4[8] = \frac{c_4[2]}{(n-1)^{1/2}} \frac{(n-2)!!}{(n-3)!!},$$

where

$$n!! = \prod_{i=0}^{\lceil n/2 \rceil - 1} (n - 2i)$$

is the double factorial of n, and $\lceil \cdot \rceil$ IS THE CEILING FUNCTION, i.e., $\lceil a \rceil$ GIVES AS OUTPUT THE SMALLEST INTEGER GREATER THAN OR EQUAL TO a. Thereby, a possible expression for $c_4[n]$, when n is even, is given by

$$c_4[n] = \frac{c_4[2]}{(n-1)^{1/2}} \frac{\prod_{i=2}^{n-2} i^{I_i}}{\prod_{i=3}^{n-3} i^{I_i^c}} = \frac{c_4[2]}{(n-1)^{1/2}} \frac{\prod_{i=2}^{n-2} i^{I_i}}{\prod_{i=2}^{n-2} i^{I_i^c}} = \frac{c_4[2]}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{I_i^*}, \qquad (5)$$

where

$$I_i = \begin{cases} 1 & \text{if } i \text{ is even} \\ 0 & \text{if } i \text{ is odd} \end{cases}$$

The indicator variables I_i^c and I_i^* are defined as

$$I_i^c = 1 - I_i = \begin{cases} 0 & \text{if } i \text{ is even} \\ 1 & \text{if } i \text{ is odd} \end{cases}$$

and

$$I_i^* = I_i - I_i^c = \begin{cases} 1 & \text{if } i \text{ is even} \\ -1 & \text{if } i \text{ is odd} \end{cases}$$

Similarly, if we evaluate the expression (4) up to $n = 9$ we obtain (SEE APPENDIX II FOR MORE DETAIL):

$$c_4[9] = \frac{c_4[2]^{-1}}{(n-1)^{1/2}} \frac{(n-2)!!}{(n-3)!!},$$

Hence, the suggested expression, when n is odd, is given by

$$c_4[n] = \frac{c_4[2]^{-1}}{(n-1)^{1/2}} \frac{\prod_{i=3}^{n-2} i^{I_i^c}}{\prod_{i=2}^{n-3} i^{I_i}} = \frac{c_4[2]^{-1}}{(n-1)^{1/2}} \frac{\prod_{i=2}^{n-2} i^{I_i^c}}{\prod_{i=2}^{n-2} i^{I_i}} = \frac{c_4[2]^{-1}}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{-I_i^*}. \tag{6}$$

The following Theorem 1 provides a new expression for the constant $c_4[n]$, and which is justified by the equations (5) and (6).

Theorem 1. For a fixed value of n, with $n \geq 4$, the function $c_4[n]$ can be expressed as

$$c_4[n] = \frac{c_4[2]^{I_n^*}}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{I_i^* I_n^*}, \quad \text{for } n \geq 4, \tag{7}$$

where I_n^* is defined by I_i^* after substituting i by n.

The proof of Theorem 1 can be seen in the Appendix I. Note that this proof uses the principle of mathematical induction to demonstrate the validity of (7). IN ADDITION, NOTE THAT (7) IS VALID FOR $n \geq 4$ DUE TO THE LIMITS OF THE PRODUCT OPERATOR OF THIS EQUATION, i.e., $\prod_{i=2}^{n-2}$ CAN BE USED IF $n \geq 4$.

3. Description of properties and discussion

We now describe some desirable properties for the proposed expression (7). First, we emphasize that the suggested expression provides the exact value of $c_4[n]$, i.e., approximations are not considered. Second, existing expressions for $c_4[n]$ cannot provide its exact value in the case of large sample sizes, since they depend on Gamma or factorial functions and such function tend to infinite as their corresponding arguments increase. However, we can observe that the suggested expression (7) does not suffer from this problem, and it is calculable for the case of large sample sizes. In addition, we observe that the suggested expression (7) is not recursive formula, since it does not depend on the previous sample size. Accordingly, it is expected that the suggested expression will be less time consuming than recursive expressions. A Monte Carlo simulation study is now carried out to compare empirically the computing time of both suggested and recursive expressions (equations (7) and (4), respectively). The simulation study is programmed with the statistical software R, and the codes are available under request. This empirical study consists of calculating the expression (7) and (4) under different sample sizes. This process is repeated 10^6 times, and the total required time (in seconds) is computed. Results derived from this empirical study can be seen in Figure 1. We observe that the proposed expression for $c_4[n]$ is less time consuming than the recursive expression, and the time difference increases as the sample size increases. For both suggested and recursive expressions, we also calculated the linear regression models between the sample size (x) and the total computing time (y), i.e,

$$y_i = \beta_0 + \beta_1 x_i + u_i,$$

where the u_i are independent and identically distributed random variables with zero mean. In Table 1, we can observe the estimation of the regression coefficients (β_1) for the proposed and recursive expressions of $c_4[n]$. For the

Table 1: Regression coefficients ($\widehat{\beta}_1$), standard errors ($\widehat{SE}[\widehat{\beta}_1]$), 95 percent confidence intervals (L and U denote the lower and upper limits, respectively), and coefficients of determination (R^2) of the linear regression models between the sample size (n) and the total computing time of the simulation study.

Expression of $c_4[n]$	$\widehat{\beta}_1$	$\widehat{SE}[\widehat{\beta}_1]$	L	U	R^2 (%)
Proposed	0.0479	0.0010	0.0458	0.0501	99.0
Recursive	0.1627	0.0010	0.1606	0.1647	99.9

proposed expression, the value of the regression coefficient is $\widehat{\beta}_1 = 0.0479$, which is clearly smaller than the value of $\widehat{\beta}_1$ for the recursive expression ($\widehat{\beta}_1 = 0.1627$). In addition, the 95 percent confidence intervals for β_1 are computed, and they are denoted as $[L, U]$, where L and U are, respectively, the lower and upper confidence limits. Such confidence intervals indicate the existence of a significant difference between the slope of both linear regression models. This issue implies that the time difference between the computing time of the proposed and recursive expressions will be greater as the sample size increases.

The commented properties suggest that the proposed expression (7) may be a convenient solution in computer programming, since it is calculable for large sample sizes and less time consuming than recursive expressions.

References

References

[1] Mitra, A. "Fundamentals of quality control and improvement", John Wiley & Sons (2016).

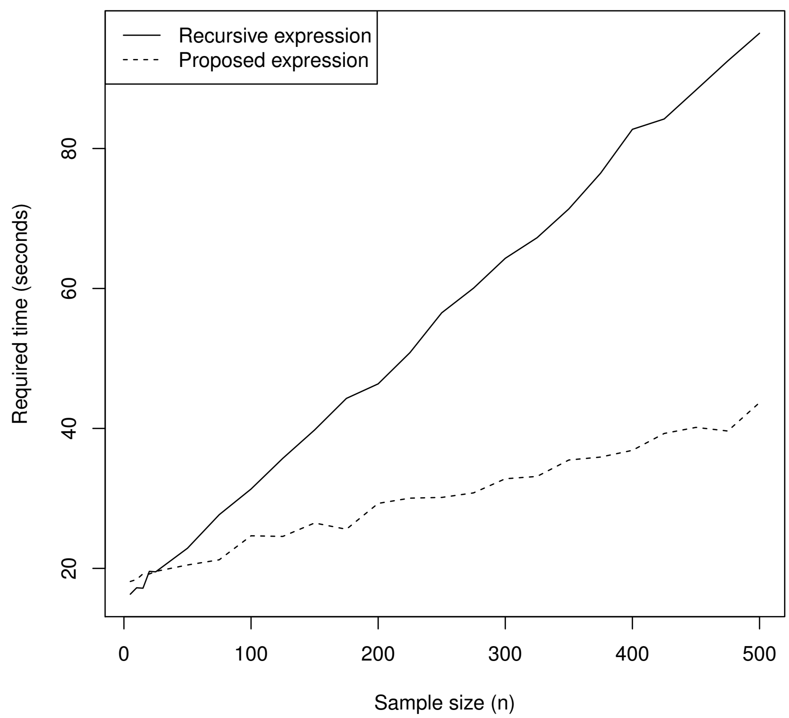

Figure 1: Total time required (in seconds) to calculate 10^6 times the constant c4 under the expressions (7) and (4) and based on several sample sizes. The X axis shows the sample sizes (n).

[2] Montgomery, D. C. “Statistical quality control (Vol. 7)”, New York: Wiley (2009).

[3] Chen, G. “The mean and standard deviation of the run length distribution of X charts when control limits are estimated”, Statistica Sinica, **7**(3), pp. 789-798 (1997).

[4] Chou, Y. M., Mason, R. L. and Young, J. C. “The SPRT control chart for standard deviation based on individual observations”, Quality Technology & Quantitative Management, **3**(3), pp. 335-345 (2006).

[5] Huang, W. H., Yeh, A. B. and Wang, H. “A control chart for the lognormal standard deviation”, Quality Technology & Quantitative Management, **15**(1), pp. 1-36 (2018).

[6] Ajadi, J. O. and Riaz, M. “New memory-type control charts for monitoring process mean and dispersion”, Scientia Iranica, Transaction E: Industrial Engineering, **24**(6), pp. 3423-3438 (2017).

[7] Diko, M.D., Chakraborti, S., Does, R.J.M.M. “Cuaranteed in-control performance of the EWMA chart for monitoring the mean”, Quality Reliability and Engineering International, **35**(4), pp.1144-1160 (2019).

[8] Chen, C. H. and Chou, C. Y. “Economic design of product and process parameters under the specified process capability value”, Quality Technology & Quantitative Management, **15**(6), pp. 686-701 (2017).

[9] Liao, M. Y. “Process capability control chart for non-normal data–evidence of on-going capability assessment”, Quality Technology & Quantitative Management, **13**(2), pp. 165-181 (2016).

[10] Polansky, A. M. and Maple, A. "Using Bayesian models to assess the capability of a manufacturing process in the presence of unobserved assignable causes", Quality Technology & Quantitative Management, **13**(2), pp. 139-164 (2016).

[11] Keshteli, R. N., Kazemzadeh, R. B., Amiri, A. and Noorossana, R. "Developing functional process capability indices for simple linear profile", Scientia Iranica, Transaction E: Industrial Engineering, **21**(3), pp. 1044-1050 (2014).

[12] Kasprikova, N. and Klufa, J. "AOQL sampling plans for inspection by variables and attributes versus the plans for inspection by attributes", Quality Technology & Quantitative Management, **12**(2), pp. 133-142 (2015).

[13] Robertson, B. L., McDonald, T., Price, C. J. and Brown, J. A. "A modification of balanced acceptance sampling", Statistics & Probability Letters, **129**, pp. 107-112 (2017).

[14] Mahmoud, M. A., Henderson, G. R., Epprecht, E. K. and Woodall, W. H. "Estimating the standard deviation in quality-control applications", Journal of Quality Technology, **42**(4), pp. 348-357 (2010).

[15] Muñoz-Rosas, J. F., Álvarez-Verdejo, E., Pérez-Aróstegui, M. N. and Gutiérrez-Gutiérrez, L. "Empirical Comparisons of X-bar Charts when Control Limits are Estimated", Quality and Reliability Engineering International, **32**(2), pp. 453-464 (2016).

[16] Chen, S. M., Liaw, J. T. and Hsu, Y. S. “Sample size determination for C_p comparisons”, Scientia Iranica, Transaction E: Industrial Engineering, **23**(6), pp. 3072-3085 (2016).

[17] Parry, G., Vendrell-Herrero, F., and Bustinza, O. F. “Using data in decision-making: Analysis from the music industry”, Strategic Change, **23**(3-4), pp.265-277 (2014).

[18] Kuczma, M. “An introduction to the theory of functional equations and inequalities: Cauchy’s equation and Jensen’s inequality”, Springer Science & Business Media (2009).

[19] Bolch, B. W. “More on unbiased estimation of the standard deviation”, The American Statistician, **22**(3), pp. 27 (1968).

[20] Cryer, J. D. and Ryan, T. P. “The Estimation of Sigma for an X Chart: $\bar{M}R/d_2$ or S/c_4?”, Journal of Quality Technology, **22**(3), pp. 187-192 (1990).

[21] Gurland, J. and Tripathi, R. C. “A simple approximation for unbiased estimation of the standard deviation”, The American Statistician, **25**(4), pp. 30-32 (1971).

[22] Team, R. C. “R: A language and environment for statistical computing”, Vienna, Austria: R Foundation for Statistical Computing (2016).

[23] Huberts, L. C., Schoonhoven, M., Goedhart, R., Diko, M. D., and Does, R. J. “The performance of control charts for large non-normally distributed datasets”, Quality and Reliability Engineering International, **34**(6), pp. 979-996 (2018).

[24] Franklin, J. "Proof in mathematics: an introduction" (1996).

[25] Hermes, H. "Introduction to mathematical logic", Springer Science & Business Media (2013).

[26] Chen, C. P. and Qi, F. "The best bounds in Wallis' inequality", Proceedings of the American Mathematical Society, pp. 397-401 (2005).

Appendix I

In this appendix we include the proof of Theorem 1. We use the principle of mathematical induction (Franklin [24], Hermes [25]) to demonstrate the validity of the proposed expression (7). For the base case, we show that the Theorem 1 holds for $n = 4$. By applying the recursive expression (4) at $n = 3$ and $n = 4$ we obtain,

$$c_4[4] = \frac{2^{1/2}}{3^{1/2} c_4[3]} = \frac{2^{1/2} 2^{1/2} c_4[2]}{3^{1/2}} = \frac{c_4[2] \times 2}{3^{1/2}}, \tag{8}$$

From (7) we also obtain (8), and (7) thus holds for $n = 4$. For the inductive step, we assume that (7) holds for some value $n \geq 4$ (induction hypothesis), and prove that $c_4[n+1]$ also holds. First, we assume that $n+1$ is even:

$$c_4[n+1] = \frac{c_4[2]^{I^*_{n+1}}}{n^{1/2}} \prod_{i=2}^{n-1} i^{I^*_i I^*_{n+1}} = \frac{c_4[2]}{n^{1/2}} \prod_{i=2}^{n-1} i^{I^*_i} = \frac{c_4[2]}{n^{1/2}} \frac{(n-1)!!}{(n-2)!!}.$$

Since $(n-1)! = (n-1)!!(n-2)!!$, we obtain:

$$c_4[n+1] = \frac{c_4[2]}{n^{1/2}} \frac{(n-1)!!^2}{(n-1)!}$$

Since $n!! = 2^{n/2}(n/2)!$ (see Chen and Qi [26]), we obtain:

$$c_4[n+1] = \frac{c_4[2]}{n^{1/2}} \frac{\left[2^{(n-1)/2}((n-1)/2)!\right]^2}{(n-1)!} = \frac{c_4[2]}{n^{1/2}} \frac{2^{n-1}((n-1)/2)!^2}{(n-1)!}.$$

From equation (3) we observe that

$$c_4[n+1] = \frac{c_4[2]}{n^{1/2}} \frac{2^{n-1}((n-1)/2)!^2}{(n-1)!},$$

which proves that (7) holds when $n+1$ is even. Second, we assume that $n+1$ is odd:

$$c_4[n+1] = \frac{c_4[2]^{I_{n+1}^*}}{n^{1/2}} \prod_{i=2}^{n-1} i^{I_i^* I_{n+1}^*} = \frac{1}{n^{1/2}c_4[2]} \prod_{i=2}^{n-1} i^{-I_i^*} = \frac{1}{n^{1/2}c_4[2]} \frac{(n-1)!!}{(n-2)!!}.$$

Since $(n-1)! = (n-1)!!(n-2)!!$, we obtain:

$$c_4[n+1] = \frac{1}{n^{1/2}c_4[2]} \frac{(n-1)!}{(n-2)!!^2}$$

Since $n-2$ is even and using the previous property of the double factorial:

$$c_4[n+1] = \frac{1}{n^{1/2}c_4[2]} \frac{(n-1)!}{\left[2^{(n-2)/2}((n-2)/2)!\right]^2} = \frac{1}{n^{1/2}c_4[2]} \frac{(n-1)!}{2^{(n-2)/2}((n-2)/2)!^2}.$$

From equation (3) we observe that

$$c_4[n+1] = \frac{1}{n^{1/2}c_4[2]} \frac{(n-1)!}{2^{(n-2)/2}((n-2)/2)!^2},$$

which proves that (7) holds when $n+1$ is odd, and this completes the proof.

Appendix II

EXPRESSIONS (5) and (6) ARE BASED ON THE EVALUATION OF THE EQUATION (4) up to $n = 9$. THIS APPENDIX CONTAINS EXPRESSIONS OF (4) WHEN THIS EQUATION IS EVALUATED AT $n = \{3, 4, \ldots, 9\}$.

$$c_4[3] = \frac{c_4[2]^{-1}}{2^{1/2}}.$$

$$c_4[4] = \frac{2^{1/2}}{3^{1/2} c_4[3]} = \frac{2^{1/2} 2^{1/2} c_4[2]}{3^{1/2}} = \frac{c_4[2]}{3^{1/2}} \frac{2}{1}.$$

$$c_4[5] = \frac{3^{1/2}}{4^{1/2} c_4[4]} = \frac{3^{1/2}}{4^{1/2}} \frac{3^{1/2}}{c_4[2]} \frac{1}{2} = \frac{c_4[2]^{-1}}{4^{1/2}} \frac{3 \times 1}{2}.$$

$$c_4[6] = \frac{4^{1/2}}{5^{1/2} c_4[5]} = \frac{4^{1/2}}{5^{1/2}} \frac{4^{1/2}}{c_4[2]^{-1}} \frac{2}{3 \times 1} = \frac{c_4[2]}{5^{1/2}} \frac{4 \times 2}{3 \times 1}.$$

$$c_4[7] = \frac{5^{1/2}}{6^{1/2} c_4[6]} = \frac{5^{1/2}}{6^{1/2}} \frac{5^{1/2}}{c_4[2]} \frac{3 \times 1}{4 \times 2} = \frac{c_4[2]^{-1}}{6^{1/2}} \frac{5 \times 3 \times 1}{4 \times 2}.$$

$$c_4[8] = \frac{6^{1/2}}{7^{1/2} c_4[7]} = \frac{6^{1/2}}{7^{1/2}} \frac{6^{1/2}}{c_4[2]^{-1}} \frac{4 \times 2}{5 \times 3 \times 1} = \frac{c_4[2]}{7^{1/2}} \frac{6 \times 4 \times 2}{5 \times 3 \times 1}.$$

$$c_4[9] = \frac{7^{1/2}}{8^{1/2} c_4[8]} = \frac{7^{1/2}}{8^{1/2}} \frac{7^{1/2}}{c_4[2]} \frac{5 \times 3 \times 1}{6 \times 4 \times 2} = \frac{c_4[2]^{-1}}{8^{1/2}} \frac{7 \times 5 \times 3 \times 1}{6 \times 4 \times 2}.$$