E. Statistical and other errors

V. Blobel
Inst. für Experimentalphysik, Hamburg

Sommersemester 2002

"There is another class of errors which are not so easy to detect and for which statistical analysis is not generally useful."

R. Bevington: Data Reduction and Analysis for the Physical Sciences

This statement is only half true.

... using talks by R. Barlow (SLUO and Durham 2002)

Other errors and correlated erros

Other errors:
systematic errors, normalization errors ...

The determination or estimation of systematic errors requires an expert knowledge of the measurement process and the physics behind.

"It requires ingenuity, experience, cunning and a fair degree of paranoia. These are well provided in particle physics. It also requires honesty and bravery, which are scarcer."

R. Barlow

... and there are "correlated" errors:
Correlated errors are **always** present, if parameters are determined in a common fit or analysis, and are described by the covariance matrix. Often correlations are neglected (by mistake).

Common practice

How to write a thesis/paper?

1 Devise cuts, get results

2 Do analysis for random errors (using e.g. Poisson statistics)

3 Make big table, alter cuts by arbitrary amounts, put in table, alter cuts, put in table, alter cuts ... until time/money/supervisors patience is exhausted.

4 Add all the variations in quadrature.

5 Quote result as "systematic error, and call it "conservative"

This give the pusillanimous researcher a comfortably large error estimate to hide behind.

This practice combines evaluation of errors with checks for mistakes, in an inappropriate way.

There is no cause to include it as contribution to the systematic error, if there is no reason to suspect a systematic effect from a particular source, and if there is no numerical evidence for a significant variation.

Example for "systematic" error determination

trigger	± 5 %
hadronic energy scale	± 4 %
electron energy scale	± 1 %
electron scattering angle ± 2 mrad	± 2 %
electron energy cut	± 6 %
lepton recognition	± 12 %
track efficiency of 98 %	± 3 %
z-vertex uncertainty	± 2 %
fiducial cut	± 4 %
structure function	± 3 %
radiative corrections	± 4 %
non-resonant background	± 4 %
branching ratio	± 3.2 %
fitting procedure	± 3 %
bin-center correction	± 2 %
Luminosity	± 1.5 %
total systematic error	18 %

Why is an uncertainty called "error"? – A proposal

The result of a measurement has to be accompanied by a quantitative statement of its **uncertainty**. Each component of uncertainty is represented by an estimated standard deviation, termed

standard uncertainty $(= +\sqrt{V})$.

The **uncertainty** of the result of a measurement generally consists of several components which may be grouped into two categories according to the method used to estimate their numerical values:

A. those which are evaluated by statistical methods,

B. those which are evaluated by other means.

A Type B evaluation is usually based on scientific judgment using all the relevant information available, which may include previous measurement data, experience with the behavior and property of releveant materials and instruments, calibration data.

The error of a measurement $(= \hat{x} - x_{\text{true}})$ in general is unknown.

Guidelines for Evaluating and Expressing the Uncertainty of Measurement Results, Technical Note 1297, by B.N. Taylor and C.E. Kuyatt, National Institute of Standards and Technology (NIST), USA

Correlated errors

Consider the formula for the transverse momentum (in GeV/c)

$$p_T = 0.3 \cdot B \cdot R$$

where

0.3 is a constant (speed of light $\times 10^{-9}$);

R is radius (in meter) of curvature from a track; for each track measured with some accuracy;

B is magnetic field (in Tesla); measured *once* with finite accuracy, but influencing every track.

No problem for one track: e.g. transverse momentum p_T has error 5 %, if we know R to 4 % and B to 3 %.

For two tracks the transverse momenta are correlated and this has to be taken into account, if anything is done with the two tracks (e.g. forming a mass).

General law of approximate "error" propagation for $\boldsymbol{x} \rightarrow \boldsymbol{y}$:

$$\boxed{\boldsymbol{V}_y = \boldsymbol{A}\boldsymbol{V}_x\boldsymbol{A}^T}$$

Transformation of errors – one track

Measured (2-dim) and transformed quantity (1-dim):

$$\text{Measured:} \quad \boldsymbol{x} = \begin{pmatrix} R_1 \\ B \end{pmatrix} \qquad \boldsymbol{V_x} = \begin{pmatrix} \sigma_{R_1}^2 & 0 \\ 0 & \sigma_B^2 \end{pmatrix}$$

$$\text{Transformed quantity:} \quad y = p_{T1} = 0.3BR_1$$

Resulting covariance "matrix"

$$\begin{aligned} \boldsymbol{V}_y &= \boldsymbol{A}\boldsymbol{V}_x\boldsymbol{A}^T \qquad \text{with} \quad \boldsymbol{A} = \begin{pmatrix} 0.3B & 0.3R_1 \end{pmatrix} \\ &= (0.3B\sigma_{R_1})^2 + (0.3R_1\sigma_B)^2 \end{aligned}$$

Transformation of "errors" – two tracks

$$\text{Measured:} \quad \boldsymbol{x} = \begin{pmatrix} R_1 \\ R_2 \\ B \end{pmatrix} \qquad \boldsymbol{V_x} = \begin{pmatrix} \sigma_{R_1}^2 & 0 & 0 \\ 0 & \sigma_{R_2}^2 & 0 \\ 0 & 0 & \sigma_B^2 \end{pmatrix}$$

$$\text{Transformed quantity:} \quad \boldsymbol{y} = \begin{pmatrix} p_{T1} \\ p_{T2} \end{pmatrix} = \begin{pmatrix} 0.3BR_1 \\ 0.3BR_2 \end{pmatrix}$$

Resulting covariance matrix

$$\begin{aligned} \boldsymbol{V}_y &= \boldsymbol{A}\boldsymbol{V}_x\boldsymbol{A}^T \qquad \text{with} \quad \boldsymbol{A} = \begin{pmatrix} 0.3B & 0 & 0.3R_1 \\ 0 & 0.3B & 0.3R_2 \end{pmatrix} \\ &= \begin{pmatrix} (0.3B\sigma_{R_1})^2 & 0 \\ 0 & (0.3B\sigma_{R_2})^2 \end{pmatrix} + \begin{pmatrix} (0.3R_1\sigma_B)^2 & (0.3\sigma_B)^2 R_1R_2 \\ (0.3\sigma_B)^2 R_1R_2 & (0.3R_2\sigma_B)^2 \end{pmatrix} \\ &= \begin{pmatrix} (0.3B\sigma_{R_1})^2 + (0.3R_1\sigma_B)^2 & (0.3\sigma_B)^2 R_1R_2 \\ (0.3\sigma_B)^2 R_1R_2 & (0.3B\sigma_{R_2})^2 + (0.3R_2\sigma_B)^2 \end{pmatrix} \end{aligned}$$

Combining random and systematic errors

Statistical and systematic errors are assumed to be independent (by definition): add in quadrature.

Consider two measurements x_1 and x_2 with

- individual (independent) random errors σ_1 and σ_2, and
- common systematic error S.

Joint covariance matrix:

$$\boldsymbol{V} = \begin{pmatrix} \sigma_1^2 + S^2 & S^2 \\ S^2 & \sigma_2^2 + S^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} + \begin{pmatrix} S^2 & S^2 \\ S^2 & S^2 \end{pmatrix}$$

The "estimation" of a combined covariance matrix may be dangerous: a mistake will produce *unexpected* (and wrong) results (see example later).

Three measurements

Consider three measurements x_1, x_2 and x_3 with

- individual independent random errors σ_1, σ_2 and σ_3, and
- common systematic error S.
- common systematic error T, shared by x_2 and x_3, but not x_1.

$$\text{cov. matrix} \qquad \boldsymbol{V} = \begin{pmatrix} \sigma_1^2 + S^2 & S^2 & S^2 \\ S^2 & \sigma_2^2 + S^2 + T^2 & S^2 + T^2 \\ S^2 & S^2 + T^2 & \sigma_3^2 + S^2 + T^2 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix} + \begin{pmatrix} S^2 & S^2 & S^2 \\ S^2 & S^2 & S^2 \\ S^2 & S^2 & S^2 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & T^2 & T^2 \\ 0 & T^2 & T^2 \end{pmatrix}$$

Example: effect of multiple scattering in material on space points measured in a track detector.

What is the "nature" of the error (or deviation)?

- B is uniform and constant, and was measured with a flux meter good to 1%.
- B is uniform and constant. It was measured with a highly accurate flux meter. However changes in geometry since then (e.g. inserting a drift chamber, opening and closing doors) will have changed that field by an amount we estimate at 1%.
- B is uniform and well measured, however it varies with time (temperature effects on power supplies) at the 1 % level.
- B is constant but not uniform: different tracks see different fields, but still obey $p_T = 0.3BR$.
- B is constant but not uniform: the same track sees different fields along it path, and $p_T = 0.3BR$ is inadequate.
- B is constant, uniform and measured accurately, but a wrong number has been typed in
- tracks loose momentum in the beam pipe so all the measurements are low
- tracks loose momentum in the drift chamber so they are not pure circular helices
- the drift chamber is badly aligned
- ...

Fitting correlated data – an example

Problem of fitting correlated data discussed by G. D'Agostini:

Paper "On the Use of the Covariance Matrix to Fit Correlated Data" (Desy 93-175), NIM A 346 (1994) 306 and `http://www-zeus.roma1.infn.it/~agostini`

Example: $X_1 = 8.0 \pm 2\%$, $X_2 = 8.5 \pm 2\%$ and common error of $\varepsilon = 10\%$ with model: true values of X_1 and X_2 are identical.

Conclusion of the author: *Systematic uncertainties can produce* (in the context of the Least Squares Method) *results that "contradict expectations."*

Fit a theoretical curve through experimental data by "minimizing the χ^2"

$$\chi^2 = \boldsymbol{\Delta}^T \boldsymbol{V}^{-1} \boldsymbol{\Delta}$$

where $\boldsymbol{V}$ is the covariance matrix and $\boldsymbol{\Delta}$ is the vector of the differences between the theoretical and the experimental values (standard method of least squares).

Numerical Example with discrepancy

Numerical example, discussed in the paper: " ...the result of two measurements, $8.0 \pm 2\%$ and $8.5 \pm 2\%$, have a $\varepsilon = 10\%$ common normalization error. Assuming that the two measurements refer to the same physical quantity, the best estimate of its true value can be obtained by fitting the points to a constant function. Minimizing χ^2 as defined above, with $\boldsymbol{V}$ estimated empirically by the data, one obtains a value of 7.87 ± 0.81, which is at least surprising[a], since the most probable result is outside the interval determined by the two measured values" ("Peelles pertinent Puzzle").

Origin of apparent discrepancy?

(D'A Paper) "*...shows that the effect is a direct consequence of the hypothesis used to estimate the empirical covariance matrix, namely the linearization on which the usual error propagation rely.*"

[a] The weigthed mean ignoring common error is $8.235 \pm 1.4\%$.

(Almost) correct treatment

Is something wrong with the principle of least squares?

(**A**): The model

$$E[x_1] = E[x_2] = \mu$$

(both single measurements are unbiased estimates of the same physical value)

$$\boldsymbol{V} = \begin{pmatrix} \sigma_1^2 + (\varepsilon\mu)^2 & (\varepsilon\mu)^2 \\ (\varepsilon\mu)^2 & \sigma_2^2 + (\varepsilon\mu)^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} + \begin{pmatrix} (\varepsilon\mu)^2 & (\varepsilon\mu)^2 \\ (\varepsilon\mu)^2 & (\varepsilon\mu)^2 \end{pmatrix}$$

(two contributions: a variance σ_1^2 and σ_2^2 for each single value, and a common normalization error, given as relative error).

Result from Least Squares:

$$8.235 \pm 1.4\% \pm 10\%$$

exactly identical to the expected result.

Approximate treatment, used by d'Agostini

(**B**): Replace $(\varepsilon\mu)$ by (εx_1) and (εx_2), respectively:

$$\boldsymbol{V} = \begin{pmatrix} \sigma_1^2 + (\varepsilon x_1)^2 & \varepsilon^2 x_1 x_2 \\ \varepsilon^2 x_1 x_2 & \sigma_2^2 + (\varepsilon x_2)^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}^2 + \begin{pmatrix} (\varepsilon x_1)^2 & \varepsilon^2 x_1 x_2 \\ \varepsilon^2 x_1 x_2 & (\varepsilon x_2)^2 \end{pmatrix}$$

Result with "empirical" covariance matrix:

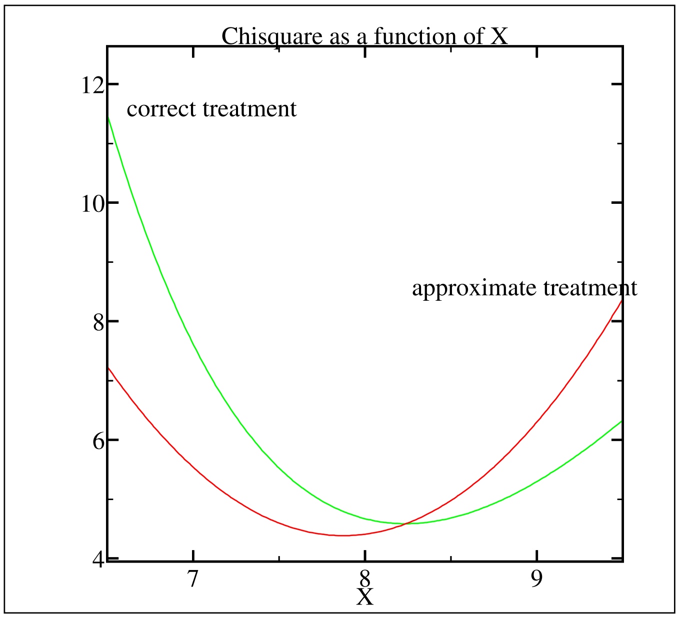
$$7.87 \pm 0.81 \qquad \rightarrow \text{ discrepancy}$$

Ansatz must be wrong, because normalization factor changed.

Geometrical explanation for difference between models (**A**) and (**B**): angle φ of principal axis of error ellipse given by $\tan 2\varphi = 2V_{12}/(V_{22} - V_{11})$:

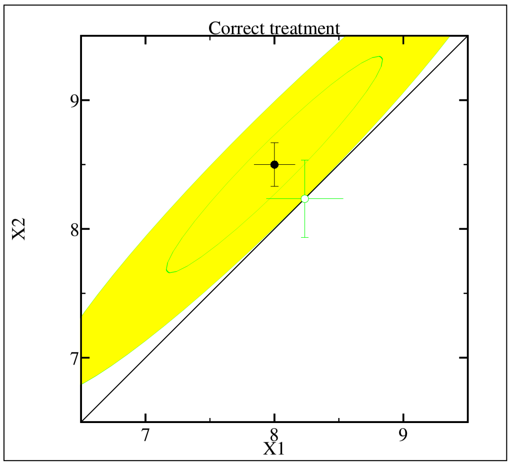
method	$\tan 2\varphi$	φ
correct treatment	-453	45.06°
approximation	-16	46.79°

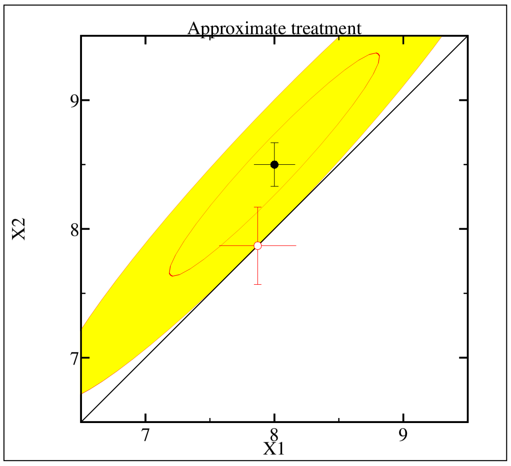
Parabolas



Ellipses

Correct treatment: axis of covariance ellipse under 45°





Approximate treatment: axis of covariance ellipse slightly tilted

The Poisson model

Is the model correct?

(**C**): Data x_i are cross sections from (Poisson distributed) event numbers n_i:

$$x_i = f \cdot n_i \qquad i = 1, 2$$

where the factor f may have relative uncertainty ε.

Optimal method for the combination of data: Maximum Likelihood based on Poisson statistic[a] with common mean value μ, with result

$$\hat{\mu} = \frac{1}{2}(n_1 + n_2) \qquad \text{(result of example would be 8.25)}$$

(The sum of two Poisson distributed numbers follows the Poisson distribution).

Average cross section: $\hat{x} = 1/2\, f \cdot (n_1 + n_2)$.

[a] Example: the case $n_1 = 1$, $n_2 = 4$. The average based on the correct Poisson distribution is simply $1/2(n_1 + n_2) = 2.5$, while the weighted average from the two separate values $n_1 = 1 \pm 1$, $n_2 = 4 \pm 2$ would be 1.6.

Diagnosis of systematic errors

Distinguish between

- errors from manifest sources (e.g uncertain calibration applied to all energy measurements)
  → determine error contribution
- unsuspected errors (e.g. change in detector performance with external conditions)
  → check for systematic effects, but usually no error contribution.

"Statistical" methods can often be used to determine "systematic errors".

For errors added in quadrature the final error is usually dominated by one or two contributions → try to reduce the large ones.

Analysis ist most effective at the planning stage of an experiment.

Manifest systematic errors

These are factors with known errors, used explicitly in the data analysis, e.g.

- conversion factors and associated errors from calibration run,
- predicted background contribution depending on paramaters used in the MC simulation, ...

Standard procedure is to repeat the analysis with e.g. modified constants, establishing the dependence of the final result on the constant.

Special case: no number quantifies the assumption, eg. different background predicted by different models (HERWIG, JETSET, ...).

Estimation method 1: take several assumptions and take standard deviation of correction factor as standard deviation of systematic error (for two values: $\sigma \approx \Delta/\sqrt{2}$; $\sigma = 0.707\Delta$.

Estimation method 2: take values under extreme assumptions, and assume $\sigma = \Delta$, divided by $\sqrt{12}$ (uniform dist.); $\sigma = 0.289\Delta$.

Unsuspected systematic errors

e.g. failure of an electronic component, new source of noise, program bug, ...

Make checks for absence of systematic effects, i.e. repeat analysis with minor differences and calculate shifts of parameters.

This is not the measurement of an effect, but a check wether an effect exists, with the hope to find no effect. Do nothing if shifts of parameters are "small" (see later).

A large shift after a minor modification would indicate the presence of an unsuspected systematic effect, which has to be treated seriously (e.g. influence of TGV on LEP energy). If it can not be *explained* it is folded into the systematic error as a last resort.

Check for systematic effects

Systematic errors which are not anticipated

Standard and sensible practice is to repeat the analysis in different forms, for example

- varying the range of the data used to extract the result,
- varying cuts applied to ensure the purity and quality of the data,
- including/excluding subsets of data taken under different experimental conditions,
- using histograms with different bin sizes,
- determining quatities ny simple counting and by fits of a paramatrized curve,

i.e. repeat the analysis with minor differences with goal.

Values obtained by different methods will disagree, even in the absence of systematic effects. If the size of such deviations may occur due to statistical fluctuations $\rightarrow$ do nothing.

Comparison of two methods of estimation - significance

Suppose the parameter a is estimated from one data sample by two different methods:

- estimate $\widehat{a}_1$ with statistical error σ_1 (method I)
- estimate $\widehat{a}_2$ with statistical error σ_2 (method II)

The difference Δ between the estimates and its variance is

$$\Delta = \widehat{a}_1 - \widehat{a}_2 \qquad {\sigma_\Delta}^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$$

In general the correlation ρ between the two estimates will be positive (estimates share some of the same data). For $\rho = +1$

$$\sigma_\Delta^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2 = (\sigma_1 - \sigma_2)^2$$

Is the difference Δ large? Determine correlation coefficient ρ and test value of Δ/σ_Δ.

Example mean of subset of the data

Estimate a_1 is the mean value of the total number N_T of data:

$$a_1 = \frac{1}{N_T}\sum_{N_T} x_i \qquad \sigma_1 = \frac{\sigma}{\sqrt{N_T}}$$

Check: using a subset of the data has the result $a_2 \pm \sigma_2$

$$a_2 = \frac{1}{N_S}\sum_{N_S} x_i \qquad \sigma_2 = \frac{\sigma}{\sqrt{N_S}}$$

Is the difference $\Delta = a_1 - a_2$ small?

The relations

$$\mathrm{cov}(a_1, a_2) = N_S \frac{\sigma}{N_T}\frac{\sigma}{N_S} \qquad \rightarrow \qquad \rho = \sigma_1/\sigma_2$$

have the result

$$\sigma_\Delta^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2 = \sigma_2^2 - \sigma_1^2 \quad \text{(i. e. subtract in quadrature)}$$

General limits on the correlation coefficient

Calculate weighted mean of the two estimates:

$$\widehat{a} = w\widehat{a}_1 + (1-w)\,\widehat{a}_2 \qquad \sigma^2 = w^2\sigma_1^2 + (1-w)^2\,\sigma_2^2 + w\,(1-w)\,\rho\sigma_1\sigma_2$$

Minimum error σ is obtained from

$$\partial\sigma^2/\partial w = 0 \rightarrow \qquad w = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

$$\sigma^2 = \frac{\sigma_1^2\sigma_2^2\left(1-\rho^2\right)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

From statistical theory there is a **Minimum Variance Bound (MVB)** σ_0^2 (see example).

For $\sigma^2 \geq \sigma_0^2$ there is an expression for ρ which is satisfied only between the two roots of the quadratic equation:

$$\rho = \frac{\sigma_0^2 \pm \sqrt{(\sigma_1^2-\sigma_0^2)(\sigma_2^2-\sigma_0^2)}}{\sigma_1\sigma_2}$$

Limits on the correlation coefficient contnd.

Allowed range: $\qquad \sigma_\Delta^{\min} \leq \sigma_\Delta \leq \sigma_\Delta^{\max}$

$$\sigma_\Delta^{\max} = \sqrt{(\sigma_1^2-\sigma_0^2)} + \sqrt{(\sigma_2^2-\sigma_0^2)} \qquad \text{(from minimum value of } \rho\text{)}$$

$$\sigma_\Delta^{\min} = \left|\sqrt{(\sigma_1^2-\sigma_0^2)} - \sqrt{(\sigma_2^2-\sigma_0^2)}\right| \qquad \text{(from maximum value of } \rho\text{)}$$

Range of ρ often narrow; if one estimator highly efficient, then MVB saturated and

$$\sigma_\Delta = \sqrt{|\sigma_1^2 - \sigma_2^2|}$$

Example: angular distribution of μ pairs in Z decays

The (normalized) probability distribution is assumed to be

$$P(x) = \frac{3}{8}\left(1 + x^2\right) + ax \qquad \text{with } x \simeq \cos\vartheta$$

where a is the asymmetry between "forward" and "backward" events. With $N = 33\,464$ events the asymetry a is determined by two methods.

- $a_1 = 0.0123 \pm 0.0055$ by counting of forward and backward events,
- $a_2 = 0.0084 \pm 0.0051$ by fitting the curve.

Does the difference of $\Delta = 0.0039$ indicate a systematic effect?

From the Minimum Variance Bound:

$$\frac{1}{\sigma_0^2} = N \int_{-1}^{+1} P(x) \left(\frac{\mathrm{d}\ln P}{\mathrm{d}a}\right)^2 \mathrm{d}x = \frac{8N}{3} \int_{-1}^{+1} \frac{x^2}{1 + x^2 + 8Ax/3}\,\mathrm{d}x \approx \frac{8N}{3}\frac{4-\pi}{2}$$

with result $\sigma_0 = 0.0051$ (equal to value σ_2 from fit).

$$\sigma_\Delta = \sqrt{\sigma_1^2 - \sigma_1^2} = 0.0021 \qquad |\Delta/\sigma_\Delta| = 0.0039/0.0021 < 2$$