Distribution of parameters

For one parameter $a$, in the fit hypothesis $y(x)=a$, we know that $a$ is normally distributed with standard deviation $\sigma_a = \Delta a$.

For independent parameters $a, b$, we can write the joint probability density function as a product:

$$f(a, b) = f(a) f(b) = \frac{1}{2\pi\sigma_a\sigma_b} e^{-\frac{1}{2}\left(\frac{(a-\mu_a)^2}{\sigma_a^2} + \frac{(b-\mu_b)^2}{\sigma_b^2}\right)}$$

We can write the term in the exponent as

$$\frac{(a-\mu_a)^2}{\sigma_a^2} + \frac{(b-\mu_b)^2}{\sigma_b^2} = (a-\mu_a, b-\mu_b)C^{-1}\begin{pmatrix} a-\mu_a \\ b-\mu_b \end{pmatrix} C = \begin{pmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{pmatrix}$$

The resulting distribution can be represented graphically.
Graphical representation (uncorrelated data)

Plot3D[
  PDF[MultinormalDistribution[{-1, 1},
      {{1, 0}, {0, 3}}], {x, y}, {x, -4, 2},
      {y, -2, 4}, PlotRange -> All,
      ColorFunction -> "TemperatureMap",
      AxesLabel -> {"a", "b", "N[μ, C]"}]]
• \( f(a,b) = \text{const.} \) is described by ellipses without tilt (uncorrelated data):

ContourPlot[
  PDF[MultinormalDistribution[{-1, 1},
    {{1, 0}, {0, 3}}, {x, y}, {x, -4, 2},
    {y, -2, 4}], PlotRange -> All,
  ColorFunction -> "TemperatureMap",
  FrameLabel -> {"a", "b", "N[\mu,C]"}]

... but this is not yet what we observed:

→ the description of the observed parameter distribution requires a rotation of the ellipse (non-zero off-diagonal terms in \( C \)!)

PHYS 6710: Nuclear and Particle Physics II
More general expression for the joint density: We know that the parameter in the one-parameter case is Normal distributed. If we fix now all other parameters, the one remaining parameter is Normal distributed:

\[ b = 0.7 = \text{const.} \]

Parameter \( a \) still Normal distributed with \( b \) fixed. Analogously for parameter \( b \), with \( a \) fixed. We are looking at conditional probabilities \( f(a|b) \). Remember: if \( x,y \) dependent,

\[ f(x, y) = f(x)f(y|x) = f(y)f(x|y) \]

We are looking for the general distribution that is Normal distributed if all but one parameters are held fixed, for all parameters. → Multivariate Normal distribution. Sound derivation: e.g., here.
The multivariate Normal distribution

- The multivariate Normal distribution for \( k \) parameters is given by

\[
f(x) = \frac{1}{(2\pi)^{k/2} \sqrt{|C|}} e^{-\frac{1}{2}(x-\mu)^T C^{-1} (x-\mu)}
\]

with the symmetric, so-called covariance matrix \( C \).

- Note that \( C_{ij} = C_{ji} \neq 0 \) for \( i \neq j \)

- As we have shown

\[
\hat{\chi}^2(\theta) = \chi^2(\hat{\theta}) + (\theta - \hat{\theta})^T C^{-1}(\theta - \hat{\theta}), \quad C^{-1} = A^T \Sigma^{-1} A
\]

we know now the distribution of parameters for the general linear regression. Also, as argued before, any chi-square from non-linear fits can be locally approximated by a quadratic form → we know that the parameters for any non-linear, non-pathological fit hypothesis follow approximately the multivariate chi-square distribution in the vicinity of the minimum.
The covariance matrix

• To finally make the connection to the “covariance”, we define:

• Let \( \mathbf{X} = (X_1, \ldots, X_n)^T \) be a random vector with \( X_i \) being random variables with finite variances, then the covariance is defined as

\[
C_{ij} = \text{Cov}(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)], \text{ with } \mu_i = E[X_i]
\]

• This can be written as outer product, \( C = E[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T] \)

\[
C = \begin{pmatrix}
E[(X_1 - \mu_1)^2] & \cdots & E[(X_1 - \mu_1)(X_n - \mu_n)] \\
\vdots & \ddots & \vdots \\
E[(X_n - \mu_n)(X_1 - \mu_1)] & \cdots & E[(X_n - \mu_n)^2]
\end{pmatrix}
\]

• By definition, the diagonal elements are the variances, \( C_{ii} = E[(X_i - \mu_i)^2] = \sigma_i^2 \) which is why \( C \) is sometimes called the variance-covariance matrix.
(continued)

- Note: if the random variables are independent, then the covariance matrix is a diagonal matrix,

\[ C = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2) \]

- The covariance matrix \( C \) is positive-semidefinite and symmetric.

- Property: with matrix \( A \), \( \text{Cov}(AX, AX) = ACov(X, X)A^T \)

- For two random variables \( X_1, X_2 \) the covariance matrix can be parameterized in terms of the “correlation coefficient” \( \rho \),

\[
\rho = \text{corr}(X_i, X_j) = \frac{\text{Cov}(X_i, X_j)}{\sigma_{X_i} \sigma_{X_j}}, \quad i \neq j \text{ and } i, j = 1, 2
\]

- Then,

\[
C = \begin{pmatrix}
\sigma_{X_1}^2 & \rho \sigma_{X_1} \sigma_{X_2} \\
\rho \sigma_{X_1} \sigma_{X_2} & \sigma_{X_2}^2
\end{pmatrix}
\]

- A non-zero \( \rho \) induces a tilt in the ellipse.
Estimate of the covariance matrix

- The quantity
  \[ Q = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^T (x_i - \bar{x}) \]
  is an unbiased estimator of the covariance matrix where \( x_i \) is the random vector in the \( i \)-th observation and
  \[ \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \]

- Proof just analogously to the case of the variance, shown at beginning of the course.

- Connection between this \( C \) and and the \( C \) appearing in the exponent of the multivariate distribution is finally made by:

- A normal-distributed random vector of independent observations with covariance matrix \( C \) is multivariate distributed with the same covariance matrix \( C \).
Connection between sample covariance and covariance in the multivariate distribution

- The maximum-likelihood estimator of the covariance matrix appearing in the multivariate distribution is given by the biased estimator of the sample covariance,

\[ Q = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^T (x_i - \bar{x}) \]

- Proof: Lengthy. Can be found here.

- This finally provides the connection between the sample covariance and its distribution. We can estimate the covariance matrix from the data and then simply plug it into the multivariate distribution, to get an estimate of the parameter distribution.
Parameter error for one parameter

• ...regardless of what values the other parameters have: obtained through *marginalization* over all other parameters. For two parameters: Show (homework!) that

\[
\int_{-\infty}^{\infty} db \, f(a, b) = f(a)
\]

where \( f(a,b) \) is the bi-variate distribution and \( f(a) \) is the Normal distribution in parameter \( a \) with variance \( \sigma_a^2 \).

• Thus, the parameter error of \( a \) is indeed given by \( \sigma_a \).

• Can we read off the parameter error directly from the ellipses of constant \( \chi^2 \)?
Parameter error from the ellipse

- **Show**: The parameter error $\sigma_a$ of $a$ (i.e., error of a parameter irrespective of what values the other parameters take), is given by the maximal extension of the error ellipse defined by

$$\chi^2(a, b, \ldots) - \chi^2(a^0, b^0, \ldots) = 1$$

i.e. the contour where chi-square has increased by 1, compared to the best “true” chi-square.

**Proof (blackboard)**: use the two conditions
- Maximal extension defined by gradient showing in $a$-direction.
- Change in chisquare is 1.

**Note**: we can find $\sigma_a$ also by:
- Searchin the $a$ for which chi-square increases by 1, optimizing at the same time all other parameters (see Minuit manual). Then, $\sigma_a = a - a^0$
- Clear from the figure
- More general than linear regression
- Serves as definition of parameter error for general, nonlinear fits.
Parameter confidence regions

- A *different* question is: What is the confidence region of parameters?

- We are not asking any more what the uncertainty of one parameter is, irrespective of the value of the others, but what the region of parameters is in which $(\hat{a}, \hat{b}, \ldots)$ lies within a certain confidence (e.g. 1-$\sigma$ confidence, 68%)

- This is answered by the following theorem: The quantity

$$ (\theta - \mu)^T C^{-1} (\theta - \mu) $$

is chi-square distributed with $m$ degrees of freedom, where $\theta$ is an $m$-dimensional random vector. Proof: by diagonalizing $C^{-1}$ and noting that the the transformation matrix is orthogonal (because $C^{-1}$ is symmetric).
(continued)

- The confidence region is then given by the region enclosed by an ellipse defined through
  \[ \Delta \chi^2 = p, \quad \text{CDF}[\chi_m^2(p)] = 1 - \alpha \]

  where a common choice is \( 1 - \alpha = 0.68 \)

- Note: for two parameters, we do not have \( \Delta \chi^2 = 1 \) any more but \( \Delta \chi^2 \approx 2.3 \)

- The confidence region is larger than the \( \Delta \chi^2 = 1 \) ellipse that was used to determine the error of one parameter.
- Only for a one-parameter fit, the parameter error and the confidence region are identical
- Note: if 1-sigma confidence is chosen then
  \[ \text{CDF}[\chi_1^2(1)] \approx 0.68 \]
Summary

Four ways to calculate the covariance matrix:

- Estimate the covariance matrix directly from the correlated data.

- Calculate from second derivative of chi-square as function of the parameters: Expand

\[ \hat{\chi}^2(\theta) = \chi^2(\hat{\theta}) + \frac{1}{2}(\theta - \hat{\theta})^T H(\theta - \hat{\theta}) + \ldots \]

where \( H \) is the Hessian matrix containing the second derivatives of the chi-square. Then, immediately:

\[ C = 2H^{-1} \]

- Calculate in linear regression: \( C^{-1} = A^T \Sigma^{-1} A \)

- Or estimate by numerical simulation, in which synthetic data is generated around real data (very similar to last exercise on sheet 2). This is called Bootstrap \( \rightarrow \) Exercises!