A Log Likelihood Fit for Extracting the D^0 Lifetime

Ilaria Manco
CID: 00938658

Abstract

The lifetime τ of the D^0 meson was extracted from a dataset of 10,000 decay time measurements by means of a parameter point estimation obtained through the maximum log-likelihood method. When the background signal was accounted for in the model, a false positive error of (1.6 ± 0.7)% due to background signal was estimated in the data sample and the average lifetime was found to be (0.410 ± 0.005) ps, in agreement with the current best measurement of (0.4101 ± 0.0015) ps [1].

1 Introduction

In high-energy physics statistical methods are often used to assess the validity of a theoretical model and find its parameters, given a large sample of data.

A common example is the estimation of the lifetime of subatomic particles. Given a large dataset of measured decay times, a theoretical model provides a predicted distribution of the measured data for which the lifetime is an unknown parameter. Hence, through a fitting process of the measured data to the predicted distribution, the parameter of interest can be inferred with some level of uncertainty [2].

2 Theory

2.1 Parameter Estimation

The decay time *t* of a subatomic particle cannot be measured directly as it is of the order of picoseconds, but it can be evaluated, assuming a known mass m, through a measurement of the particle's momentum $\boldsymbol{p}$ and the vector $\boldsymbol{x}$ connecting the spatial points of its production and decay [3]:

$$t = m\,\frac{\boldsymbol{x}\cdot\boldsymbol{p}}{p^2} \tag{1}$$

and it is expected to follow an exponential decay function $f^t(t)$ given by

$$f^t(t) = \begin{cases} 0 & if\ t < 0 \\ 1/\tau\ e^{-t/\tau} & if\ t \geq 0, \end{cases} \tag{2}$$

where τ is the average lifetime of the particle.

Each measurement of *t*, however, has an associated uncertainty due to the error in the measurement of the particle's position and therefore follows a probability density function (PDF) $f^m(t)$ that results from the convolution of the theoretical decay function (eq. 2) with a Gaussian of width σ, representing the limited detector resolution:

$$f^m(t) = \frac{1}{2\tau}\exp\left(\frac{\sigma^2}{2\,\tau^2} - \frac{t}{\tau}\right) erfc\left(\frac{1}{\sqrt{2}}\left(\frac{\sigma}{\tau} - \frac{t}{\sigma}\right)\right), \tag{3}$$

where σ is the measurement error.

Since the measurements are independent, the values of *t* and σ are fixed and the combined probability distribution of *N* measurements is the product of the individual probabilities. The likelihood that the distribution fits the data is then given by the likelihood function $\mathcal{L}(\tau)$:

$$\mathcal{L}(\tau) = \prod_{i=1}^{N} f^m(t_i, \sigma_i), \tag{4}$$

where (t_i, σ_i) are the pairs of measured decay times and their error and *N* the total number of measurements in the fit.

Evaluating the unknown parameter $\hat{\tau}$ at which the $\mathcal{L}(\tau)$ is maximised therefore allows us to obtain an estimate of the true value of the mean life of the particle.

For practical reasons, the negative logarithm of the likelihood function

$$-\ln\mathcal{L}(\tau) = -\sum_{i=1}^{N}\ln\big(f^m(t_i, \sigma_i)\big) \tag{5}$$

is considered instead, giving the *negative log-likelihood*, which is minimised at the same value of the parameter τ as $-\mathcal{L}(\tau)$.

In order to improve the model, a nuisance parameter can be added to account for the level of background in the measurements. The probability density function $f_{bkg}^{m}(t)$ of the background is given by the convolution of a delta function with a Gaussian [3]:

$$f_{bkg}^{m}(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{t^2}{\sigma^2}\right) \qquad (6)$$

and including this in the model gives a multivariate distribution as the PDF for the decay time, dependent on both the lifetime τ and the fraction of signal in the sample a:

$$f(t) = a\, f^{m}(t) + (1-a) f_{bkg}^{m}(t) \qquad (7)$$

The *negative log-likelihood* $-\ln \mathcal{L}(\tau, a)$ hence becomes dependent on two parameters.

2.2 Parameter Uncertainty

According to the Central Limit Theorem, in the limit of a large number of data, the *log-likelihood* function can be approximated by a Gaussian, which in the multi-dimensional case this will be a multivariate distribution.

Neglecting terms of order higher than the second allows to find the statistical uncertainty $\widehat{\sigma}$ on the parameter (as an approximation) from the second derivative of the *negative log-likelihood*:

$$\frac{1}{\widehat{\sigma}^2} = \frac{d^2 \ln \mathcal{L}}{d\tau^2}. \qquad (8)$$

For a multivariate distribution (assuming the Gaussian approximation is valid), the same result can be obtained from the covariance matrix $\boldsymbol{V}$ of the likelihood function, which, in the asymptotic limit, is given by the inverse of the Hessian matrix $\boldsymbol{H}^{-1}$ [4], whose elements are

$$H_{ij} = \frac{\partial^2 \ln \mathcal{L}}{\partial \tau_i\, \partial a_j}, \qquad (9)$$

evaluated at $\tau_i = \hat{\tau}$ and $a_j = \hat{a}$.

The diagonal elements of the matrix give the variance of the parameters, and the off-diagonal elements the correlation $\rho\widehat{\sigma}_\tau\widehat{\sigma}_a$ between the parameters:

$$\boldsymbol{V}(\hat{\tau}, \hat{a}) = \begin{pmatrix} \hat{\sigma}_\tau & \rho\hat{\sigma}_\tau\hat{\sigma}_a \\ \widehat{\rho\sigma}_\tau\hat{\sigma}_a & \hat{\sigma}_a \end{pmatrix}, \qquad (10)$$

where ρ is the correlation coefficient.

A numerically similar result can also be obtained, in the asymptotic limit [4], by finding the value of the parameter(s) at which the *log-likelihood* varies by 0.5 with respect to the maximum, corresponding to a variation of $1\widehat{\sigma}$ (estimated standard deviation) (fig. 1).

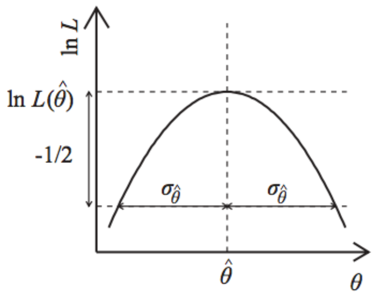

Figure 1 For an (approximately) Gaussian log-likelihood, the uncertainty in the parameter θ is given by a shift of the likelihood function of 0.5 in absolute units compared to its maximum (or minimum for a NLL) value [4].

Hence, in the one-dimensional case, an approximation of the error can be derived by solving for τ

$$\Delta(-\ln \mathcal{L}) = 0.5, \qquad (11)$$

where $\Delta(-\ln \mathcal{L}) = \ln \mathcal{L}(\tau) - \ln \mathcal{L}_{\min}$ and the two solutions give the parameter's uncertainty $\widehat{\sigma} = \tau - \hat{\tau}$ in the negative and positive interval respectively.

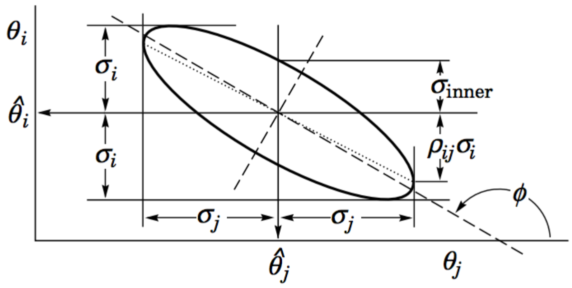

Figure 2 Standard error ellipse with negative correlation. [5]

Equivalently, when the parameters to estimate are more than one (as in the case of the PDF that includes the background signal), the uncertainty intervals can be found from the points on the distribution's surface that define the contour of $\Delta(-\ln \mathcal{L}) = 0.5$, also called the standard error ellipse, illustrated in fig.(2) [5].

The extreme limits of this contour ellipse, correspond to the boundary of a confidence interval of 68% (one standard deviation) and the parameters' uncertainties are given by the projections onto the axes of the furthest points on the ellipse from the minimum.

3 Method

3.1 Estimation without background

A set of 10,000 pairs of (t_i, σ_i) measurements of the D^0 meson decay was used as the data sample for the point estimation of the best-fit parameter $\hat{\tau}$.

A histogram of the data with its fit function superimposed for different hypothetical values of the parameter τ and measurement uncertainty σ is presented below.

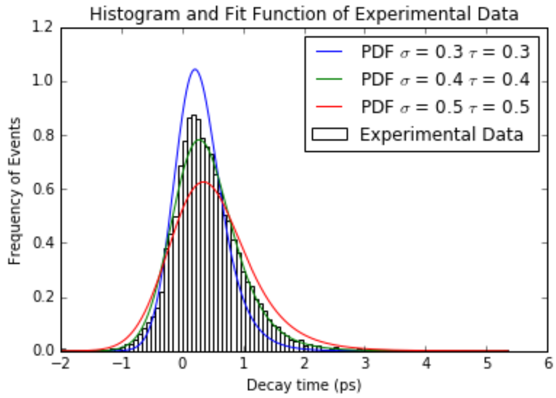

Figure 3 Histogram of the decay times in the data sample with a plot of the PDF $f(t)$ for three different values of the measurement uncertainty σ and lifetime τ. By inspection, a rough estimate of the best fit parameter $\hat{\tau} \approx 0.4$ ps can be inferred.

The integral of the probability density function was calculated for different values of σ_i and τ to ensure that this was normalised and independent of the parameter τ.

A one-dimensional minimisation of the *negative log-likelihood* function (NLL) (eq.5) was implemented using an inverse parabolic interpolation: three initial points, *a*, *b*, *c*, within an interval bracketing the minimum, were chosen and a second-order Lagrange polynomial was fitted through them; the minimum *d* of the interpolating parabola was found and the process was iterated using the three points out of *a*, *b*, *c* and *d* at which the NLL was the lowest, until the requirement for convergence was satisfied. This was chosen to happen when a new iteration gave the same minimum as the previous within a tolerance interval of 10^{-5} ps, as this corresponded to the precision of the data.

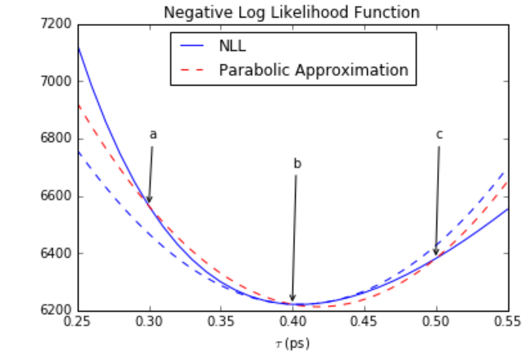

Figure 4 NLL plot around the minimum. The dashed lines are the parabolic approximation in the first (red) and last (blue) iteration of the parabolic minimization respectively. The initial points were chosen to lie in the interval (a, c) bracketing the minimum.

The initial interval was determined through a rough estimate of the minimum from the NLL plot (fig.4) and the minimiser was tested on $f(x) = \cosh(x)$ to ensure a correct functionality.

3.1.1 Accuracy

The estimate of the parameter's uncertainty in the 1D minimisation was found following the two approaches introduced in section 2.2.

A finite difference approximation of the second derivative of the NLL was used to calculate the curvature of the function at the minimum and thus obtain another estimate of the standard error; this was also compared to the curvature of the parabola at the minimum to validate the parabolic approximation used in the optimisation.

Equivalently, an additional estimation of the statistical error was found through the $\Delta(-\ln \mathcal{L})$ method. For this purpose, the Newton-Raphson method was implemented to find the two roots of the function in eq. 11 providing the value of the standard deviation.

3.2 Estimation with background signal

A similar approach was adopted in the parameter optimisation for the bivariate distribution (eq. 7).

Three multidimensional minimisers were developed, using the gradient method, the

Newton method and the quasi-Newton method in order to find the values of τ and a minimising the NLL.
A rough estimate of the parameters to optimise was obtained through a contour plot of the likelihood function, as shown in fig. 5.

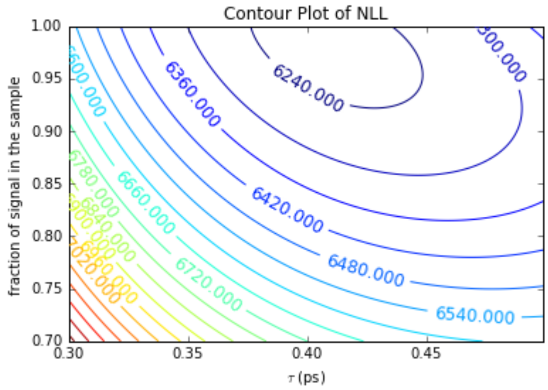

Figure 5 Contour plot of the 2D NLL with parameters τ (x-axis) and a (y-axis). The approximate minimum point can be observed to lie within 0.35 < τ < 0.45 ps and 0.90 < a < 1.0.

Finite differences of the first and second derivatives of the NLL were used to calculate the gradient $\underline{\nabla}(-\ln \mathcal{L})$ and the Hessian matrix $\boldsymbol{H}$ of the function, used in the gradient (and quasi-Newton) and Newton method respectively.
An approximation of the inverse of the Hessian, $\boldsymbol{G}$, was found through the Davidon-Fletcher-Power algorithm [3] (refer to .py file) and used in the implementation of the quasi-Newton method.
The efficiency, stability and convergence of each algorithm were manually tested for the specific case of the optimisation here discussed by measuring their running time and varying the starting values of the parameters in the first iteration. As a result of this, an initial point corresponding to $\tau =$ 0.4 and a = 0.9 was chosen as the most suitable for all three methods and a step size $\alpha = 10^{-5}$ was used in the gradient and quasi-Newton optimisation.

3.2.1 Accuracy

The statistical error in the 2D parameter estimation was evaluated using a similar procedure to the 1D case: instead of the second derivative of the NLL, the inverse of the covariance matrix was evaluated at the minimum point and the diagonal elements, representing the variance of the parameters, were found.
As an additional estimation of the uncertainties, the Newton method was first used by keeping one of the variables (τ or a) fixed at their best estimate value, while varying the other, thus obtaining $\hat{\sigma}_{\text{inner}}$ for each of the parameters (see fig. 2); then this root-find procedure was applied again for varying values of a in order to obtain the standard error ellipse and thus the positive and negative uncertainty interval.

4 Results and Discussion

4.1 Estimation without background

Through the 1D minimisation of the *negative log-likelihood* fit, the average lifetime τ of the D^0 meson was found as the most optimal point of the model's parameter space and was estimated to be (0.405 ± 0.005) ps when the false background readings were neglected in the model.
The standard error on the parameter obtained by changing the value of the NLL by 0.5 (fig. 6) and through the curvature both delivered the same numerical result, although slightly asymmetric confidence intervals resulted from the first method, as expected. Within the accuracy of 10^{-15} ps, however, the parabolic approximation holds and the *negative log-likelihood* can be considered parabolic around its minimum. Hence, the Gaussian limit can also be correctly assumed in this case.

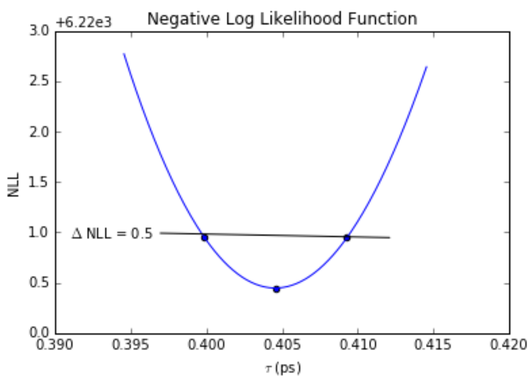

Figure 6 NLL around the minimum (middle point). The two external points correspond to the negative and positive value at which the NLL increases by 0.5.

The relationship between the number of measurements N included in the fit and the statistical error associated with the parameter was also investigated and the standard deviation was found to be proportional to $1/\sqrt{N}$, as expected. A linear fit was used to obtain the constant of proportionality (fig. 7) and this was found to be 0.44 and 0.43 using the $\Delta(-\ln \mathcal{L})$ and the curvature method for finding the standard error respectively, as explained above. In analogy with the Gaussian distribution, the constant of proportionality thus found can then be identified with the distribution's width (the uncertainty introduced by the detector in this case) [2].

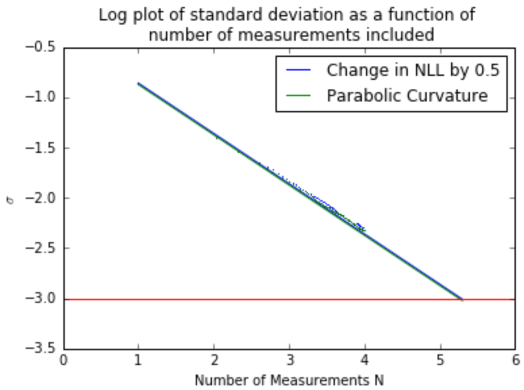

Figure 7 $\log_{10} \hat{\sigma}_\tau$ as a function of the number of measurements N included in the fit (on a $\log_{10}$ scale). The two lines represent the linear fits of the numerical values (points) obtained for the standard deviation of τ with the $\Delta(-\ln \mathcal{L})$ (blue) and the curvature (green) method. The red line corresponds to an accuracy of 10^{-15} ps and the intersection between the blue or the green line gives the number of measurements required to get a statistical error of that order.

From this fit, a prediction of the number of data points needed to increase the accuracy of the result to 10^{-3} ps was extracted and this was found to be between 180,320 and 191,736 measurements. However, due to the approximate nature of the uncertainties' estimates, a considerable statistical error is associated with this value.

4.2 Estimation with background

Adding to the model the fraction of background signal by including the nuisance parameter a (fraction of signal in the sample), resulted in a more accurate

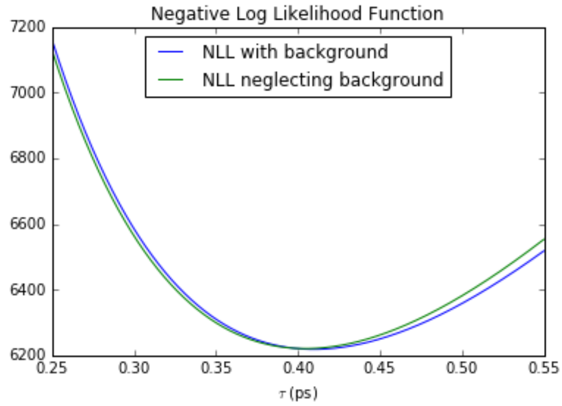

Figure 8 NLL of the PDF when neglecting the background (green) and when including this in the model (blue). The difference in the best fit parameter $\hat{\tau}$ between the two is of 0.0051 ps.

estimation (fig. 8), yielding a value of $\hat{\tau} = 0.410$ ps, with a fractional difference of 1% from the current best value[1] and a shift of 0.0051 ps compared to the estimate obtained in the first estimation (sec. 4.1).

The 2D minimisation results also provided an estimate for the fraction of signal in the sample, which was found to be $0.9837^{-0.0087}_{+0.0083}$.

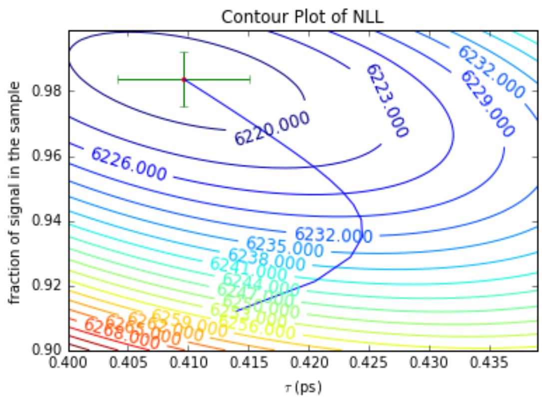

Figure 9 Two-dimensional NLL around the minimum $\tau = 0.410$ ps and $a = 0.9837$ (red point) with its error (green bars). The blue curve perpendicular to the contour lines represents the gradient method minimisation.

The presence of an additional parameter, however, despite reducing the effect of the systematic error introduced by the background, increased the statistical uncertainty, resulting in a final estimation of the lifetime of $0.4096^{-0.0054}_{+0.0056}$ and an uncertainty interval which no longer exhibits symmetry and therefore more accurately describes the PDF, which is indeed characterised by a positive skewness (fig. 11).

The different estimates of the parameters' uncertainties obtained through the $\Delta(-\ln\,\mathcal{L}) = 0.5$ contour (by fixing one the parameters at their best estimate first and then varying it throughout the width of the entire error ellipse) and through the diagonal elements of the covariance matrix are summarised in Table 1.

As expected, an underestimation of the error resulted from keeping one of the parameters fixed at the optimal point, as this only allows to find $\hat{\sigma}_{inner}$ (see fig. 2) along the axes and not the maximum value on the confidence interval's boundary. This can therefore only be correctly evaluated (within the approximation limit) by numerically finding the ellipse's extrema (fig. 10), which, as reported in the table, agree with the uncertainty obtained from the covariance matrix.

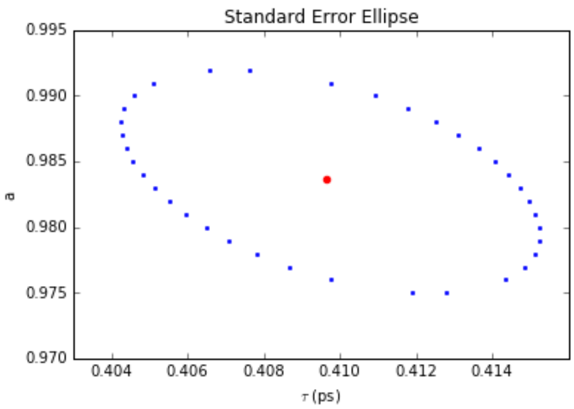

Figure 10 Points on the standard error ellipse found through the Newton method (blue). The red point is the NLL minimum.

Hence, as in the one-parameter case, the local Gaussian approximation around the minimum point of the NLL holds and hence yields a reasonably accurate parameter estimation for our model.

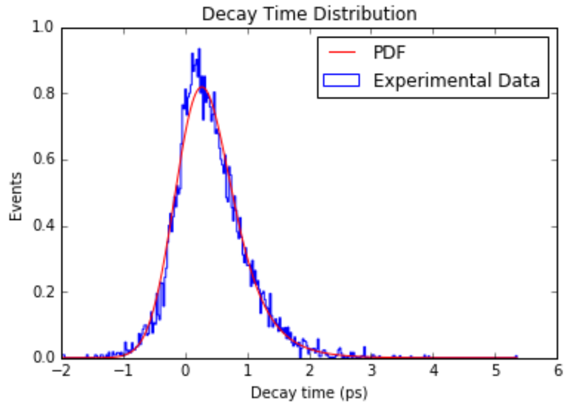

Figure 11 Histogram of the decay times in the data sample with a plot of the PDF $\boldsymbol{f(t)}$ for the best fit parameter $\boldsymbol{\tau}$ found through the minimisation of the 2D NLL (see fig. 3).

A final test that the model for which the best-fit parameters were found was consistent with the data was carried out by plotting the PDF of the decay time using the real value of τ and a over the normalised histogram of the dataset.

The good agreement between the observed decay time distribution and the theoretical fit function hence supports the use of this model and the evaluation its parameters here reported.

Parameter	**Uncertainty Estimate $\hat{\sigma}$**		
	Δ(NLL) = 0.5	Ellipse Extrema	Covariance Matrix
Lifetime τ	- 0.0047 + 0.0049	- 0.0054 + 0.0056	0.0055
Signal fraction a	- 0.0076 + 0.0074	- 0.0087 + 0.0083	0.0086

Table 1 Summary of the uncertainties estimates on the parameters τ and a of the NLL. The Δ(NLL) = 0.5 and ellipse extrema methods provide a negative and positive confidence interval, while the standard deviation found through the covariance matrix yields a symmetric uncertainty interval. All values are in units of picoseconds.

5 Conclusion

The unbinned maximum likelihood method was followed in order to extract the average lifetime of the D^0 meson from a dataset of 10,000 measurements of its decay time. This was estimated to be *(0.410 ± 0.005) ps* when taking into account a background signal of *(1.6 ± 0.7)%* in the sample, which was determined as an additional parameter.

The model was found to agree with the world-average value of τ within its error [1], and several steps were identified that could represent a further improvement: the model could be expanded by adding further nuisance parameters such as the detector resolution or its efficiency in order to investigate the contribution to the signal detection due to the source of uncertainty in the measurements and subsequently reduce the systematic error; a goodness-of-fit assessment could be carried out to fully validate the model and determine how accurately it describes the data; a Monte Carlo simulation could be implemented to generate simulated data samples according to the model's fit function and further investigate the validity of the model [6].

REFERENCES

[1] C. Patrignani et al. (Particle Data Group). Charmed Mesons. *Chin. Phys. C*. 2016. 40: 100001.

[2] L. Lista. *Statistical Methods for Data Analysis in Particle Physics*. Lecture Notes in Physics, vol. 909. Cham: Springer International Publishing AG. 2016

[3] Y. Uchida and E. van Sebille. *Computational Physics – Project B1 Script.* Imperial College London. 2016.

[4] O. Behnke, K. Kröninger, G. Schott, T. Schörner-Sadenius (eds). *Data Analysis in High Energy Physics, A Practical Guide to Statistical Methods*. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA. 2013

[5] C. Patrignani et al. (Particle Data Group). Statistics. *Chin. Phys. C*. 2016. 40: 100001. [Online] Available from: http://www-pdg.lbl.gov/2016/reviews/rpp2016-rev-statistics.pdf [Accessed 16th December 2016]

[6] T. Aaltonen et al. Measurement of the B_c^- meson lifetime in the decay $B_c^- \rightarrow J/\psi\pi^-$. *Physical Review.* 2013. D 87: 011101(R). [Online] Available from: DOI: https://doi.org/10.1103/PhysRevD.87.011101 [Accessed 16th December 2016]