L3 Note 2196

Numerical Construction of Likelihood Distributions and the Propagation of Errors

J. Swain[†,1] **and L. Taylor**[†,2]

December 12, 1997

Abstract

The standard method for the propagation of errors, based on a Taylor series expansion, is approximate and frequently inadequate for realistic problems. A simple and generic technique is described in which the likelihood is constructed numerically, thereby greatly facilitating the propagation of errors.

To appear in Nucl. Instrum. Meth.

† Department of Physics, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA
1 Electronic mail: John.Swain@cern.ch
2 Electronic mail: Lucas.Taylor@cern.ch

1 Introduction

Traditionally, errors on derived quantities have been determined analytically by performing a Taylor series expansion about the central values. This method is adequate for many simple problems but often fails in more realistic situations for a number of different reasons, as described in section 2.

In section 3 we describe a rigorous and generally applicable numerical technique for the propagation of errors, based on the numerical construction of the likelihood for the derived quantity using pseudo-random numbers. Sections 4 and 5 illustrate the use of this technique with practical examples from High Energy Physics analyses.

While the method proposed has almost certainly been used by others, it has received little attention in the literature. This is perhaps due to the fact that only in the last few years has it been practical to generate large numbers of pseudo-random numbers in order to solve a problem which was traditionally done by hand.

2 Analytic Method for the Propagation of Errors

In general, physical quantities are not known with infinite precision but are described in terms of likelihood distributions (also called probability or error distributions). To be specific, consider a quantity f which depends on some quantities x_i $(i = 1, \ldots, n)$ in a known way. The uncertainty on f, which depends on the uncertainties δx_i on the measured values x_i^0, is usually derived by the standard method for the propagation of errors which relies on a Taylor series expansion of f:

$$f(x_i^0 + \delta x_i) = f(x_i^0) + \sum_{i=1}^{n} \delta x_i \left.\frac{\partial f}{\partial x_i}\right|_{x_i = x_i^0} + \cdots \tag{1}$$

where the ellipsis denotes higher order terms in δx_i. These are neglected to yield the familiar expression for the variance of f:

$$\sigma_f^2 = \sum_{ij} \sigma_{ij} \left.\left(\frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \right)\right|_{\substack{x_i = x_i^0; \\ x_j = x_j^0}} \tag{2}$$

where σ_{ij} denotes the error matrix. For the results of this method to be valid, a number of requirements should be satisfied.

Requirement 1: The likelihood distributions of x_i are Gaussian.

The standard method for the propagation of errors implicitly assumes that the errors on the x_i are Gaussian. In general, however, a measured quantity is described by an arbitrarily shaped likelihood distribution. Statistical and systematic errors on quantities x_i may be according to other common likelihood distributions (Poisson, binomial, *etc.*), or another function which may itself depend on measured quantities with associated errors.

Requirement 2: The likelihood distribution of f is Gaussian.

The standard method for propagation of errors yields a single number σ_f which is interpreted as the RMS of the Gaussian likelihood distribution of f. It is often the case that the likelihood distribution of f is not Gaussian, nor even symmetric, even if all the errors on x_i are Gaussian. In such cases the interpretation of σ_f is unclear. A familiar example occurs in tracking detectors where the inverse transverse momentum, $1/p_{\mathrm{T}}$, has a Gaussian error whereas the derived quantity, $f(1/p_{\mathrm{T}}) = p_{\mathrm{T}}$, does not and is manifestly asymmetric.

Requirement 3: The required derivatives are calculable.

The evaluation of the derivatives, $\partial f/\partial x_i$, may prove to be difficult in practice if f is a complicated function of many parameters. If the dependence of f on x_i were determined numerically, for example by running a physics Monte Carlo program, it may be difficult or impossible to determine the required derivatives. Moreover, the derivatives may themselves have significant uncertainties due to limited Monte Carlo statistics and systematic errors from uncertainties in the values of physical parameters used by the Monte Carlo program.

Requirement 4: Higher terms of the Taylor expansion are negligible.

It is assumed that quadratic and even higher order terms in the Taylor expansion of Eq. 1 are negligible. This approximation is not always valid, for example if the first derivatives of f at x_i are zero or small compared to higher derivatives or if the errors are not small relative to the x_i.

3 Numerical Method for the Propagation of Errors

If one or more of the above conditions is not satisfied, we advocate the use of a numerical method for the propagation of errors, as described below. Even if all the conditions are satisfied in a given situation, the numerical method provides a valuable cross-check. Moreover, if the situation becomes progressively more complicated as more sources of error are identified, it is easy to extend the numerical method, whereas the standard method may ultimately fail one of the above assumptions.

The numerical method is based on the familiar statistical concept of performing many hypothetical measurements of the same quantity and determining the error on that quantity from the spread of the values. It may be summarised as follows:

> *Consider a function $f(x_i)$ where the independent variables x_i have uncertainties which are described by likelihood functions $L(x_i)$. The corresponding likelihood distribution $L(f)$ may be described with arbitrary precision by a sufficiently large set of values, $\{f(x_i')\}$, where the input set of values, $\{x_i'\}$, are chosen randomly according to the $L(x_i)$.*

While the standard method for the propagation of errors based on the Taylor expansion of Eq. 1 assumes a single δx_i, which is equated to the standard deviation of a Gaussian, the numerical method maps the full spectrum of errors δx_i and therefore does not require it to be Gaussian. Moreover, no higher order terms are neglected.

Practically one implements the numerical method as follows:

1. Define the likelihood functions, $L(x_i)$, for the independent variables, x_i.
   For example, in the case of uncorrelated Gaussian errors one would specify a set of central values and standard deviations. For correlated Gaussian errors one would specify a set of central values and a covariance matrix. In a more general case one might define the likelihood as a multi-dimensional function represented by a smooth parametrisation, a binned multi-dimensional histogram, or as a set of discrete values derived, for example, from a Monte Carlo program.

2. Repeat the following steps, (a) and (b), n times, where n is a large number:

   (a) Choose a set of values $\{x_i'\}$ randomly according to the likelihood functions, $L(x_i)$.
   For example, in the case of uncorrelated Gaussian errors one might use the CERNLIB [1] Fortran subroutine `RANNOR`. For correlated Gaussian errors one could use the NAGLIB [2] Fortran subroutine `G05EAF` or a suitably transformed set of uncorrelated random numbers [3].

(b) Evaluate the function $f(x'_i)$ and store the result, for example in a histogram or an array.

3. Normalise the integral of $L(f)$ to unity.

4. Ensure that n is sufficiently large, for example by verifying that $L(f)$ does not change significantly for different starting random number seeds.

Care should be taken to avoid numerical problems associated with the computing hardware and software used [4, page 659]. For example, if n is large then the accumulation of $f(x'_i)$ values in a histogram may be subject to numerical imprecisions, or a poorly designed random number generator may start to show periodic behaviour. The optimisers of some compilers may even treat multiple invocations of a random number function, for example in a Fortran `DO` loop, as a single invocation and move it outside the loop thereby producing erroneous results. Caveats such as these, however, apply to most software and can be avoided easily by minimal precautions such as the use of extended precision variables and low levels of compiler optimisation.

4 Example: Measurement of the W Boson Mass

W mass measurements from hadron colliders [5] are typically based on fits to distributions of transverse mass, $m_T = \sqrt{2p_T^\ell p_T^\nu(1-\cos\phi^{\ell\nu})}$, where p_T^ℓ and p_T^ν denote the transverse momenta of the charged lepton (e or μ) and the neutrino respectively coming from the leptonic W decay, and $\phi^{\ell\nu}$ denotes the angle between the charged lepton and the neutrino in the transverse plane. The charged lepton is typically measured in a tracking chamber with an error which is approximately Gaussian in $1/p_T^\ell$ while p_T^ν, which is inferred from the missing energy, has an approximately Gaussian error. The opening angle error is typically relatively small.

Consider a single typical event with $p_T^\ell = 40\,\mathrm{GeV}$, $p_T^\nu = 35\,\mathrm{GeV}$, $\phi^{\ell\nu} = 2\,\mathrm{rad}$, which corresponds to $m_T = 63.0\,\mathrm{GeV}$. Assuming that the errors on p_T^ℓ and p_T^ν are uncorrelated, then the standard method for propagation of errors using Eq. 2 yields

$$\frac{\sigma^2_{m_T}}{m_T^2} = \left(\frac{1}{2p_T^\ell}\right)^2 \sigma^2_{p_T^\ell} + \left(\frac{1}{2p_T^\nu}\right)^2 \sigma^2_{p_T^\nu} \tag{3}$$

$$= \left(\frac{p_T^\ell}{2}\right)^2 \sigma^2_{1/p_T^\ell} + \left(\frac{1}{2p_T^\nu}\right)^2 \sigma^2_{p_T^\nu} \tag{4}$$

$$\tag{5}$$

For example, given $\sigma_{1/p_T^\ell} = 0.005\,\mathrm{GeV}^{-1}$ and $\sigma_{p_T^\nu}/p_T^\nu = 10\%$ yields $\sigma_{m_T} = 7.0\,\mathrm{GeV}$.

The likelihood distribution of m_T is not in reality a Gaussian. Fig. 1 (solid line) shows the m_T likelihood distribution obtained by propagating the uncertainties on p_T^ℓ and p_T^ν numerically. A clear asymmetry is seen, in contrast to the naive method for propagation of errors which, by construction, yields a Gaussian likelihood distribution (dashed line). The shift in the peak is $\sim 2\,\mathrm{GeV}$ which is significant compared to typical errors on the W mass of $O(0.1)\,\mathrm{GeV}$.

5 Example: τ neutrino mass and mixing

Constraints on the tau neutrino mass and its mixing with a hypothetical fourth lepton generation have been derived [6] by considering the dependence of tau branching fractions on the mass m_{ν_τ} and the Cabibbo-like mixing angle θ_L (or more naturally $\sin^2\theta_L$). The theoretical predictions are

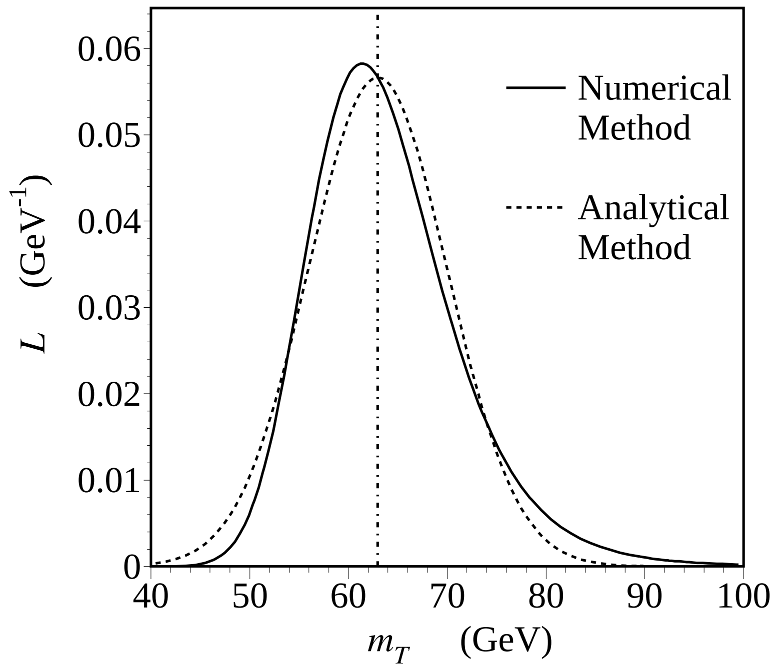

Figure 1: Example likelihood distribution of transverse mass showing the difference between the numerical and analytic approaches described in the text. Both curves are normalised to unit area.

compared with the experimental measurements for the following decays[1]: $\tau^- \to \mathrm{e}^- \bar{\nu}_{\mathrm{e}} \nu_\tau$, $\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$, $\tau^- \to \pi^- \nu_\tau$, and $\tau^- \to \mathrm{K}^- \nu_\tau$.

The experimental measurements of the branching ratios B_i^{expt} $(i = \mathrm{e}, \mu, \pi, \mathrm{K})$ are uncorrelated and have Gaussian errors. If the theoretical predictions, B_i^{theory}, were perfectly known, then the likelihood for a particular choice of m_{ν_τ} and $\sin^2\theta_L$ would be[2]

$$L(m_{\nu_\tau}, \sin^2\theta_L | B_{\mathrm{e}}^{\mathrm{expt}}, B_\mu^{\mathrm{expt}}, B_\pi^{\mathrm{expt}}, B_{\mathrm{K}}^{\mathrm{expt}}) = \prod_i \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(- \left(B_i^{\mathrm{theory}} - B_i^{\mathrm{expt}} \right)^2 \Big/ 2\sigma_i^2 \right), \qquad (6)$$

where the σ_i are the errors on the B_i^{expt}. The predictions for B_i^{theory}, however, depend in turn on experimentally measured quantities with errors.

The theoretical predictions for the branching fractions B_ℓ for the decays $\tau^- \to \ell^- \bar{\nu}_\ell \nu_\tau$, with $\ell = \mathrm{e}/\mu$, are given by [6]:

$$\begin{aligned} B_\ell^{\mathrm{theory}} &= \left(\frac{G_{\mathrm{F}}^2 m_\tau^5}{192\pi^3} \right) \tau_\tau (1 - \sin^2\theta_{\mathrm{L}}) \\ &\quad \times \left[1 - 8x - 12x^2 \ln x + 8x^3 - x^4 - 8y(1-x)^3 + \cdots \right] \\ &\quad \times \left[\left(1 - \frac{\alpha(m_\tau)}{2\pi} \left(\pi^2 - \frac{25}{4} \right) \right) \left(1 + \frac{3}{5} \frac{m_\tau^2}{m_{\mathrm{W}}^2} + \cdots \right) \right] \end{aligned} \qquad (7)$$

[1]Henceforth we denote the branching ratios for these processes as $B_{\mathrm{e}}, B_\mu, B_\pi, B_{\mathrm{K}}$ respectively; B_ℓ denotes either B_{e} or B_μ while B_{h} denotes either B_π or B_{K}.

[2]The CLEO measurement of the τ mass was used to further constrain m_{ν_τ}. From an analysis of $\tau^+\tau^- \to (\pi^+ n\pi^0 \bar{\nu}_\tau)$ $(\pi^- m\pi^0 \nu_\tau)$ events (with $n \le 2, m \le 2, 1 \le n+m \le 3$), CLEO determined the τ mass to be $m_\tau = (1777.8 \pm 0.7 \pm 1.7) + [m_{\nu_\tau}(MeV)]^2/1400MeV$ [7]. The likelihood for the CLEO and BES measurements to agree, as a function of m_{ν_τ} is included in the global likelihood. This does not affect the conclusions of this section but merely reinforces them.

where G_F is the Fermi constant, m_τ and τ_τ are the τ mass and lifetime, $x = m_\ell^2/m_\tau^2$, m_ℓ is the charged lepton mass, $y = m_{\nu_\tau}^2/m_\tau^2$, m_W is the W mass, $\alpha(m_\tau)$ is the renormalised fine-structure constant at the τ mass scale, and each ellipsis denotes neglected higher order terms. The first term in brackets allows for phase-space while the second term in brackets allows for radiative corrections [8, 9, 10, 11]. Similarly, the branching fractions for the decays $\tau^- \to \mathrm{h}^-\nu_\tau$, with $\mathrm{h} = \pi/\mathrm{K}$, are given by

$$\begin{aligned} B_\mathrm{h}^\mathrm{theory} &= \left(\frac{G_\mathrm{F}^2 m_\tau^3}{16\pi}\right)\tau_\tau f_\mathrm{h}^2 |V_{\alpha\beta}|^2 (1-\sin^2\theta_\mathrm{L}) \\ &\times \left[(1-x)^2\left(1-y\left(\frac{2+x-y}{1-x}\right)\right)\sqrt{1-y\left(\frac{2+2x-y}{(1-x)^2}\right)}\right] \\ &\times \left[1+\frac{2\alpha}{\pi}\ln\left(\frac{m_\mathrm{Z}}{m_\tau}\right)+\cdots\right] \end{aligned} \tag{8}$$

where $x = m_\mathrm{h}^2/m_\tau^2$, m_h is the hadron mass, $y = m_{\nu_\tau}^2/m_\tau^2$, f_h are the hadronic form factors, and $V_{\alpha\beta}$ are the CKM matrix elements, V_ud and V_us, for π^- and K^- respectively. The ellipsis represents terms, estimated to be $\mathcal{O}(\pm 0.01)$ [12], which are neither explicitly treated nor implicitly absorbed into G_F, $f_\pi|V_\mathrm{ud}|$, or $f_\mathrm{K}|V_\mathrm{us}|$.

The uncertainties on the B_i^theory depend on the errors on the values of: G_F, τ_τ, m_e, m_μ, m_π, m_K, m_W, and m_Z [13]; m_τ from the BES measurement at threshold [14]; $f_\pi|V_\mathrm{ud}|$ and $f_\mathrm{K}|V_\mathrm{us}|$ [12, and references therein]; and the estimated theoretical uncertainty due to (neglected) higher order radiative corrections.

If the standard method for the propagation of errors were to be applied, one would calculate the theoretical errors according to Eq. 2, by differentiation of Eqs. 7 and 8, and add them in quadrature to the experimental errors on the B_i to obtain σ_i. This approach is problematic for a number of reasons:

- the input errors are not necessarily Gaussian (for example the theoretical uncertainty on the neglected higher order terms);
- the uncertainty on B_i^theory is non-Gaussian, as may be seen immediately from just the m_τ^5 and m_τ^3 dependences of B_ℓ^theory and $B_\mathrm{h}^\mathrm{theory}$ respectively;
- many rather lengthy derivative calculations are required;
- there is no *a priori* guarantee that the neglect of higher order terms in the Taylor expansion is a reasonable approximation;
- the four B_i^theory predictions depend on many common input parameters such that the four σ_i cannot be treated as independent errors.

The numerical procedure described in section 3 avoids all of these problems. A large ensemble of values of B_i^theory is created by choosing values for G_F, τ_τ, etc. according to their errors and then evaluating $B_\mathrm{e}^\mathrm{theory}$, B_μ^theory, B_π^theory, and $B_\mathrm{K}^\mathrm{theory}$ according to Eqs. 7 and 8. The likelihood is calculated according to Eq. 6 with σ_i taken to be the error on $B_i^\mathrm{expt.}$ only. The full likelihood, allowing for the errors on $B_i^\mathrm{expt.}$, B_i^theory, and all correlations is then obtained from the normalised sum of the likelihoods for the full ensemble. To be specific, the following steps are carried out:

1. define the likelihood functions, $L(x_i)$, for the independent variables, x_i;
2. create a 2D histogram of m_{ν_τ} vs. $\sin^2\theta_\mathrm{L}$;
3. repeat the following steps, (a) and (b), n times:

(a) choose a set of values $\{x'_i\}$ randomly according to the likelihood functions, $L(x_i)$.

(b) for each bin in the histogram choose m_{ν_τ} and $\sin^2\theta_{\rm L}$ at the centre of the bin and then evaluate $L(m_{\nu_\tau}, \sin^2\theta_L | B_{\rm e}^{\rm expt}, B_\mu^{\rm expt}, B_\pi^{\rm expt}, B_{\rm K}^{\rm expt})$, according to Eqs. 6, 7, and 8, and add the value of L to the contents of the bin;

4. normalise the histogram to unity to obtain $L(m_{\nu_\tau}, \sin^2\theta_L)$.

Figure 2(a) shows the contours of the two dimensional likelihood distribution, $L(m_{\nu_\tau}, \sin^2\theta_L)$ which correspond to the 90% and 95% confidence levels. No evidence is seen for a non-zero neutrino mass, nor for mixing. By integration of the two dimensional likelihood over all values of $\sin^2\theta_{\rm L}$ we obtain the one-dimensional likelihood for m_{ν_τ}, as shown by the solid line of figure 2(b), which yields upper limits of $m_{\nu_\tau} < 42(48)\,{\rm MeV}$ at the 90(95)% confidence levels. The solid line of figure 2(c) shows the one-dimensional likelihood distribution for $\sin^2\theta_{\rm L}$, integrated over all values of m_{ν_τ}, from which we derive the upper limits: $\sin^2\theta_{\rm L} < 0.014(0.017)$ at the 90(95)% confidence levels.

6 Conclusions

The standard method for the propagation of errors, based on a Taylor series expansion, is approximate and frequently inadequate for realistic problems. In particular, it assumes that the errors on the independent quantities are Gaussian, that the error on the derived quantity is Gaussian, that the required derivatives are calculable, and that higher order terms in the Taylor expansion are negligible.

A numerical method for the propagation of errors is described which makes no such assumptions, provides exact results with arbitrary precision, and is straightforward to implement even for complicated problems.

Realistic examples illustrating this numerical technique are described. The interpretation of constraints on neutrino masses, with either a *Classical* or a *Bayesian* approach [15], has received much attention in the literature [13, 16]. In the example described herein, a flat Bayesian prior distribution has been implicitly assumed by sampling m_{ν_τ} and $\sin^2\theta_{\rm L}$ from a histogram with uniform bins. Similarly, negative neutrino masses are excluded by the choice of the histogram range. While such choices are a matter of discussion, it should be emphasised that they are in no way imposed by the use of the numerical algorithm for the propagation of errors, which is in fact generally applicable.

Acknowledgements

We would like to acknowledge many colleagues in the ARGUS and L3 collaborations for valuable discussions, in particular Bernhard Spaan. This work was supported by the US National Science Foundation.

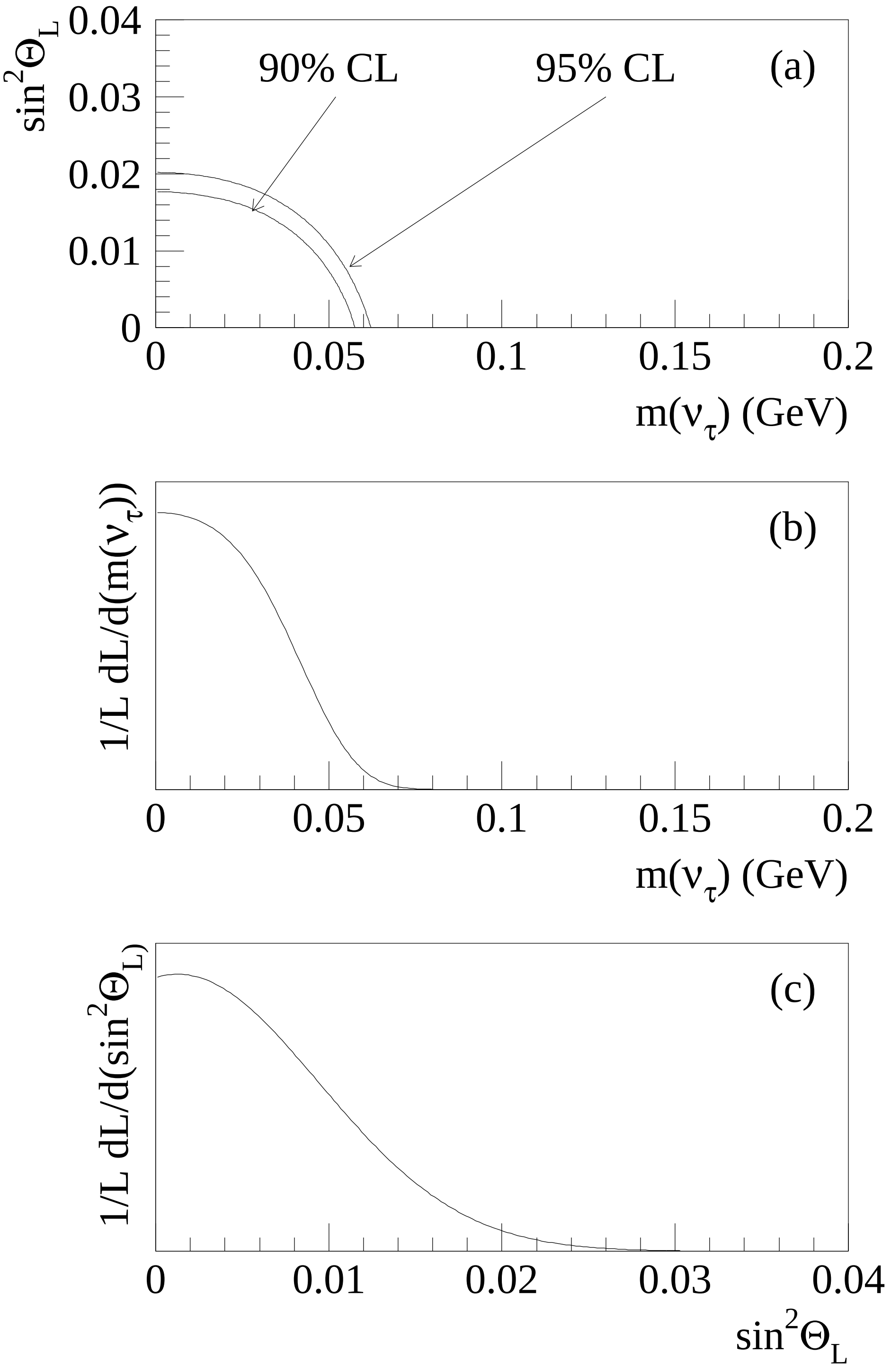

Figure 2: Likelihood distributions for all τ decay channels combined, for (a) $\sin^2\theta_{\rm L}$ *vs.* m_{ν_τ}, (b) m_{ν_τ}, integrated over $\sin^2\theta_{\rm L}$, and (c) $\sin^2\theta_{\rm L}$, integrated over m_{ν_τ}.

References

[1] See: `http://wwwinfo.cern.ch/asd/index.html`.

[2] See, for example: `http://wwwcn.cern.ch/asdoc/WWW/naglib/nagnew.html`.

[3] As B. Spaan has pointed out, it is straightforward to generate a set of random values $\vec{x}'$ according to an arbitrary covariance matrix V. First determine the eigenvalues λ_i of the covariance matrix and the corresponding eigenvectors $\vec{x}_i$. Construct a rotation matrix R from the eigenvectors where each column corresponds to an eigenvector. Then, generate a vector $\vec{r}$ with i uncorrelated elements each chosen randomly from the Gaussian distribution: $\vec{r}_i = \left[\exp\left(-x^2/2\lambda_i\right)\right]/\sqrt{2\pi\lambda_i}$. Obtain the set of correlated random numbers $\vec{x}'$ by rotation: $\vec{x}' = R \cdot \vec{r}$.

[4] F. James, Monte Carlo Theory and Practice, in *Experimental Techniques in High-Energy Nuclear and Particle Physics*, edited by T. Ferbel, World Scientific, Singapore, 1991.

[5] L. Taylor, Properties of the W Boson, in *Proceedings of the XVIIth International Conference on Physics in Collision*, edited by H. Heath, Singapore, 1997, World Scientific, in press.

[6] J. Swain and L. Taylor, Phys. Rev. **D 55**, 1 R (1997).

[7] R. Balest et al., Phys. Rev. **D47**, R3671 (1993), The neutrino mass dependence quoted in this paper is in error due to a typographical oversight (A. Weinstein, private communication). This does not, however, affect any other numbers in the paper.

[8] S. M. Berman, Phys. Rev. **112**, 267 (1958).

[9] T. Kinoshita and A. Sirlin, Phys. Rev. **113**, 1652 (1959).

[10] A. Sirlin, Rev. Mod. Phys. **50**, 573 (1978).

[11] W. J. Marciano and A. Sirlin, Phys. Rev. Lett. **61**, 1815 (1988).

[12] W. J. Marciano, Phys. Rev. **D45**, R 721 (1992).

[13] Particle Data Group, R. M. Barnett *et al.*, Phys. Rev. **D54**, 1 (1996).

[14] J. Bai, Phys. Rev. **D53**, 20 (1996).

[15] W. Eadie, D. Drijard, F. James, M. Roos, and B. Sadoulet, *Statistical Methods in Experimental Physics*, North-Holland, Amsterdam, The Netherlands, 1971.

[16] F. James and M. Roos, Phys. Rev. **D 44**, 299 (1991).