CDF/MEMO/STATISTICS/PUBLIC/5305
Version 1.00
June 24, 2005

A Convolution Method for Folding Systematic Uncertainties into Likelihood Functions

Luc Demortier
Laboratory of Experimental High-Energy Physics
The Rockefeller University

Abstract

We describe a general convolution method for incorporating systematic uncertainties into likelihood functions for analyses that seek to extract upper limits or two-sided intervals on some parameter of interest. The method is derived as an approximation to the correct Bayesian approach and therefore inherits the latter's statistical interpretation and to some degree its mathematical properties. In particular, the form of the convolution kernel is fully determined, in contrast with commonly used approaches that rely on intuition and guesswork. A simple example is analyzed in detail.

1 Introduction

High-energy physics analyses are often required to deal with a significant number of so-called nuisance parameters, i.e. parameters that are of no particular interest and yet must be specified in order to extract physics information from the data. Examples include background contaminations, tracking efficiencies, jet energy scales, initial and final state radiation, parton distribution functions, etc. Uncertainties on nuisance parameters are usually referred to as "systematic uncertainties." In the following we restrict our attention to analyses that are based on a likelihood function $\mathcal{L}(\theta, \nu \,|\, x)$, where θ labels the physics parameter, ν the nuisance parameter, and x the observed data; all three of these quantities can be vectors. The problem is then to eliminate the dependence of $\mathcal{L}$ on ν, without ignoring the effect that uncertainties on ν have on inferences about θ. Two common methods for solving this problem are profiling and marginalization.

In the profiling method [1], the likelihood function is maximized with respect to the nuisance parameter(s) at each physics parameter value of interest. This technique is often applied when the nuisance parameter is a background contribution to the

mean of a Poisson probability. In this case the dependence of the likelihood on the nuisance parameter is analytically simple, making the maximization relatively easy to implement numerically. Unfortunately this method cannot be applied to nuisance parameters with more complex dependencies, such as the jet energy scale. Another disadvantage of the profiling method is that it does not take proper account of the uncertainty on the nuisance parameter. Although corrections for this are available in the statistical literature, they are not widely known in our field and are usually neglected, regardless of whether this is justified.

In the marginalization method [2], the likelihood is integrated over the nuisance parameter(s), similarly to what would be done in a Bayesian context with a noninformative, constant nuisance prior. In HEP applications however, there is often external information available about the nuisance parameter. Assuming that this information can be represented as a Gaussian constraint with known mean and width, the Bayesian approach would then base inference about the physics parameter θ on the following function:

$$\int d\nu \; \mathcal{L}(\theta, \nu \,|\, x) \; \frac{e^{-\frac{1}{2}\left(\frac{\nu-\nu_0}{\Delta\nu}\right)^2}}{\sqrt{2\pi}\,\Delta\nu}. \tag{1.1}$$

As already pointed out, the dependence of $\mathcal{L}$ on ν is often of a very complex, nonanalytical nature, which makes the above integral intractable. An alternative is to replace it by a convolution over the physics parameter:

$$\int d\theta' \; \mathcal{L}(\theta', \nu_0 \,|\, x) \; \frac{e^{-\frac{1}{2}\left(\frac{\theta-\theta'}{\sigma}\right)^2}}{\sqrt{2\pi}\,\sigma}, \tag{1.2}$$

where ν_0 is the nominal value of the nuisance parameter and σ is some function of θ, θ', ν_0, and $\Delta\nu$. The problem now is to determine the form of this function.

The aim of this note is to solve this problem by reformulating the convolution method as an approximation to the exact (but numerically intractable) Bayesian procedure. This approach will yield a definite functional form for the width σ. At the same time, it will benefit to some extent from general features of the Bayesian method, such as its straightforward statistical interpretation and its well-understood mathematical behavior.

The derivation of the correct convolution integral is given in section 2. Section 3 describes techniques that can be used in numerical applications of the method. A simple example is worked out in detail in section 4. Section 5 contains some conclusions.

2 Derivation of the Convolution Method

The Bayesian solution to the inference problem described in the introduction starts with the computation of the posterior density:

$$p(\theta, \nu \,|\, x) \;=\; \frac{\mathcal{L}(\theta, \nu \,|\, x)\, \pi(\theta, \nu)}{m(x)}, \tag{2.1}$$

where $\pi(\theta, \nu)$ is the prior probability density for θ and ν, and $m(x)$ is the so-called predictive density:

$$m(x) \;=\; \int d\theta \int d\nu \; \mathcal{L}(\theta, \nu \,|\, x)\, \pi(\theta, \nu). \tag{2.2}$$

Elimination of the nuisance parameter ν is achieved by marginalization:

$$p(\theta \,|\, x) \;=\; \int d\nu \; p(\theta, \nu \,|\, x) \;=\; \int d\nu \; p(\theta \,|\, \nu, x)\; p(\nu \,|\, x), \tag{2.3}$$

where the second equality follows from the definition of conditional probability. Next, we write an expression for the nuisance posterior, $p(\nu \,|\, x)$:

$$p(\nu \,|\, x) \;=\; \int d\theta \; p(\theta, \nu \,|\, x) \;=\; \int d\theta \; \frac{\mathcal{L}(\theta, \nu \,|\, x)\, \pi(\theta, \nu)}{m(x)}. \tag{2.4}$$

Therefore, again using the definition of conditional probability:

$$p(\theta \,|\, \nu, x) \;=\; \frac{p(\theta, \nu \,|\, x)}{p(\nu \,|\, x)} \;=\; \frac{\mathcal{L}(\theta, \nu \,|\, x)\; \pi(\theta, \nu)}{\displaystyle\int d\theta' \; \mathcal{L}(\theta', \nu \,|\, x)\; \pi(\theta', \nu)}. \tag{2.5}$$

It is usually reasonable to assume that prior information about θ is independent of that about ν, so that the prior factorizes: $\pi(\theta, \nu) = \pi(\theta)\; \pi(\nu)$. Furthermore, since θ is the parameter of interest, it is often given a noninformative prior such as $\pi(\theta) = 1$. With these assumptions, the above equation becomes:

$$p(\theta \,|\, \nu, x) \;=\; \mathcal{L}^{\star}(\theta \,|\, \nu, x), \tag{2.6}$$

where $\mathcal{L}^{\star}(\theta \,|\, \nu, x)$ is the likelihood function normalized with respect to θ for each value of ν:

$$\mathcal{L}^{\star}(\theta \,|\, \nu, x) \;=\; \frac{\mathcal{L}(\theta, \nu \,|\, x)}{\displaystyle\int d\theta' \mathcal{L}(\theta', \nu \,|\, x)}. \tag{2.7}$$

Plugging equation (2.6) into (2.3) yields:

$$p(\theta \,|\, x) \;=\; \int d\nu \; \mathcal{L}^{\star}(\theta \,|\, \nu, x)\; p(\nu \,|\, x). \tag{2.8}$$

Finally, assume that:

$$p(\nu \,|\, x) \;\approx\; \pi(\nu), \tag{2.9}$$

so that:

$$p(\theta \,|\, x) \;\approx\; \int d\nu \; \mathcal{L}^{\star}(\theta \,|\, \nu, x)\; \pi(\nu). \tag{2.10}$$

The assumption of equation (2.9) is usually somewhat conservative, in the sense that if there is no conflict between the data and the priors, then the mean of $p(\nu \,|\, x)$ will agree with the mean of $\pi(\nu)$, but $\pi(\nu)$ will be *wider* than $p(\nu \,|\, x)$, since the posterior incorporates data information that is not available at the prior level. Therefore, smearing the likelihood with $\pi(\nu)$ instead of $p(\nu \,|\, x)$ in equation (2.10) tends to exaggerate the

effect of the uncertainty on the nuisance parameter ν. The assumption that there is no conflict between data and priors should generally be valid if the θ prior is noninformative and if the ν prior is itself derived from data, albeit in an independent, subsidiary sample. Under certain conditions, equation (2.10) can be shown to be exact.[3]

Often, the only prior information available about the nuisance parameter ν is its mean $\nu_\circ$ and standard deviation $\Delta\nu$, and a Gaussian representation is then assumed. Equation (2.10) therefore takes the form:

$$p(\theta \,|\, x) \;\approx\; \int d\nu \; \mathcal{L}^\star(\theta \,|\, \nu, x) \; \frac{e^{-\frac{1}{2}\left(\frac{\nu-\nu_\circ}{\Delta\nu}\right)^2}}{\sqrt{2\pi}\;\Delta\nu}. \tag{2.11}$$

Although the integration range in this equation was left unspecified, the Gaussian factor makes it safe to assume that the integrand will be negligible for ν values more than a few $\Delta\nu$'s away from ν_0. This is not only useful for numerical calculations, but also allows us to keep track of dominant contributions as further approximations are introduced. We therefore write:

$$p(\theta \,|\, x) \;\approx\; \int_{\nu_0-n\Delta\nu}^{\nu_0+n\Delta\nu} d\nu \; \mathcal{L}^\star(\theta \,|\, \nu, x) \; \frac{e^{-\frac{1}{2}\left(\frac{\nu-\nu_\circ}{\Delta\nu}\right)^2}}{\sqrt{2\pi}\;\Delta\nu}, \tag{2.12}$$

where n depends on the desired accuracy of the calculation and on one's confidence in the model for large values of the nuisance parameter; a reasonable value would be 3 or larger. As described in the introduction, we now wish to transform the above integral into a Gaussian convolution over θ, with a width σ to be determined as a function of $\Delta\nu$, ν_0, and θ. To achieve this we need to relate an arbitrary shift in the nuisance parameter ν to a shift in the parameter of interest θ. This can be done with the help of a shift function $\rho_\theta(\epsilon)$, defined implicitly by the equation:

$$P(\theta \,|\, \nu_\circ + \epsilon, x) \;=\; P(\theta + \rho_\theta(\epsilon) \,|\, \nu_\circ, x), \tag{2.13}$$

where:

$$P(\theta \,|\, \nu, x) \;\equiv\; \int_{-\infty}^{\theta} d\theta' \; \mathcal{L}^\star(\theta' \,|\, \nu, x). \tag{2.14}$$

The shift function can be constructed graphically, as illustrated in Figure 1. This graphical construction is particularly useful when $\mathcal{L}^\star(\theta \,|\, \nu, x)$ is obtained in histogram form. If the nuisance parameter only affects the location or the width of the likelihood, then the shift function will be linear in θ. Distortions of higher moments of the likelihood lead to non-linearities. Some examples of shift functions are shown in Figure 2.

Differentiating equation (2.13) with respect to θ leads to:

$$\mathcal{L}^\star(\theta \,|\, \nu_\circ + \epsilon, x) \;=\; \mathcal{L}^\star(\theta + \rho_\theta(\epsilon) \,|\, \nu_\circ, x) \; \left|1 + \frac{d\rho_\theta(\epsilon)}{d\theta}\right|, \tag{2.15}$$

so that, after changing the integration variable ν to $\epsilon \equiv \nu - \nu_0$, equation (2.12) can be

rewritten as:

$$p(\theta \,|\, x) \;\approx\; \int_{-n\Delta\nu}^{+n\Delta\nu} d\epsilon\; \mathcal{L}^\star(\theta \,|\, \nu_\circ + \epsilon, x)\; \frac{e^{-\frac{1}{2}\left(\frac{\epsilon}{\Delta\nu}\right)^2}}{\sqrt{2\pi}\,\Delta\nu}, \tag{2.16}$$

$$\approx\; \int_{-n\Delta\nu}^{+n\Delta\nu} d\epsilon\; \mathcal{L}^\star(\theta + \rho_\theta(\epsilon) \,|\, \nu_\circ, x)\; \left|1 + \frac{d\rho_\theta(\epsilon)}{d\theta}\right|\; \frac{e^{-\frac{1}{2}\left(\frac{\epsilon}{\Delta\nu}\right)^2}}{\sqrt{2\pi}\,\Delta\nu}. \tag{2.17}$$

Next, assuming that $\rho_\theta(\epsilon)$ is one-to-one in ϵ for all values of θ, we perform the change of variable

$$\epsilon \to \eta: \quad \eta \equiv \theta + \rho_\theta(\epsilon). \tag{2.18}$$

Without loss of generality we then assume that $\rho_\theta(\epsilon)$ *increases* with ϵ, so that:

$$p(\theta \,|\, x) \;\approx\; \int_{\theta+\rho_\theta(-n\Delta\nu)}^{\theta+\rho_\theta(n\Delta\nu)} d\eta\; \mathcal{L}^\star(\eta \,|\, \nu_\circ, x)\; \frac{e^{-\frac{1}{2}\left[\frac{\rho_\theta^{-1}(\eta-\theta)}{\Delta\nu}\right]^2}}{\sqrt{2\pi}\,\Delta\nu}\; \left|\frac{1 + \frac{d\rho_\theta(\epsilon)}{d\theta}}{\frac{d\rho_\theta(\epsilon)}{d\epsilon}}\right|_{\epsilon=\rho_\theta^{-1}(\eta-\theta)}, \tag{2.19}$$

where ρ_θ^{-1} is the inverse of the function $\rho_\theta(\epsilon)$ with respect to its ϵ argument:

$$\rho_\theta^{-1}(z) \;=\; \epsilon \quad\Leftrightarrow\quad \rho_\theta(\epsilon) \;=\; z. \tag{2.20}$$

Define σ by:

$$\sigma(\theta, \eta) \;=\; \frac{\eta - \theta}{\rho_\theta^{-1}(\eta - \theta)}\; \Delta\nu. \tag{2.21}$$

To write the Jacobian factor in the integrand of equation (2.19) in terms of σ, we first rewrite (2.21) as:

$$\eta - \theta \;=\; \rho_\theta\Big(\frac{\eta - \theta}{\sigma(\theta, \eta)}\,\Delta\nu\Big), \tag{2.22}$$

and differentiate with respect to θ:

$$-1 \;=\; \frac{\partial\rho_\theta(\epsilon)}{\partial\theta}\bigg|_{\epsilon=(\eta-\theta)\frac{\Delta\nu}{\sigma}} \;-\; \frac{\partial\rho_\theta(\epsilon)}{\partial\epsilon}\bigg|_{\epsilon=(\eta-\theta)\frac{\Delta\nu}{\sigma}}\; \left[\frac{\Delta\nu}{\sigma} \;+\; \frac{\eta-\theta}{\sigma^2}\,\Delta\nu\,\frac{\partial\sigma(\theta,\eta)}{\partial\theta}\right]. \tag{2.23}$$

The components of this equation can be rearranged as follows:

$$\left.\frac{1 + \frac{\partial\rho_\theta(\epsilon)}{\partial\theta}}{\frac{\partial\rho_\theta(\epsilon)}{\partial\epsilon}}\right|_{\epsilon=\rho_\theta^{-1}(\eta-\theta)} \;=\; \frac{\Delta\nu}{\sigma(\theta,\eta)}\left[1 \;+\; \frac{\eta-\theta}{\sigma(\theta,\eta)}\,\frac{\partial\sigma(\theta,\eta)}{\partial\theta}\right]. \tag{2.24}$$

With the help of (2.21) and (2.24), the integral (2.19) becomes:

$$p(\theta \,|\, x) \;\approx\; \int_{\theta+\rho_\theta(-n\Delta\nu)}^{\theta+\rho_\theta(n\Delta\nu)} d\eta\; \mathcal{L}^\star(\eta \,|\, \nu_\circ, x)\; \frac{e^{-\frac{1}{2}\left[\frac{\eta-\theta}{\sigma(\theta,\eta)}\right]^2}}{\sqrt{2\pi}\,\sigma(\theta,\eta)}\; \left[1 \;+\; \frac{\eta-\theta}{\sigma(\theta,\eta)}\,\frac{\partial\sigma(\theta,\eta)}{\partial\theta}\right]. \tag{2.25}$$

This is our final result: it has the form of a convolution with a Gaussian kernel, but with a non-trivial Jacobian factor in the integrand.

3 Practical Implementation

In most applications one would proceed in one of two ways for calculating the convolution integral (2.25): by quadrature or by Monte Carlo. Both methods require some type of approximation to parameterize the dependence of the shift function $\rho_\theta(\epsilon)$ on ϵ. A relatively simple approach is to first calculate the likelihood at a discrete number of nuisance parameter values, say $\nu = \nu_0 + m\Delta\nu$, with $m = 0, \pm 1, \pm 2, \pm 3, \ldots$ One can then extract the shift function $\rho_\theta(m\Delta\nu)$ versus θ from these calculations, and use standard interpolation or extrapolation techniques to estimate $\rho_\theta(\epsilon)$ at ϵ values different from $m\Delta\nu$. The integration methods described below all need $\rho_\theta(\epsilon)$ for ϵ between $-n\Delta\nu$ and $+n\Delta\nu$, so that the largest $|m|$ value should be less than n. Furthermore, as extrapolation is generally less reliable than interpolation, this largest $|m|$ should not be too much smaller than n.

The following sections discuss in some detail the quadrature and Monte Carlo approaches. The interpolation techniques presented here are all of the polynomial type, but this is clearly not a requirement of the method. Depending on how many support points are available, one could also use rational function or cubic spline interpolations.[4]

3.1 Quadrature methods

Quadrature methods obtain an integral such as (2.25) by cleverly combining several evaluations of the integrand. Many algorithms for doing this, such as the trapezoidal rule, Simpson's rule, Romberg integration, and Gaussian quadratures, are described in great detail in reference [4]. Here we limit ourselves to a few remarks about evaluating the integrand of (2.25).

Assuming that we have determined $\rho_\theta(m\Delta\nu)$ as a function of θ and for one or more values of m, the width $\sigma(\theta, \eta)$ at $\eta = \theta + \rho_\theta(m\Delta\nu)$ can be obtained by direct application of definition (2.21):

$$\sigma(\theta, \theta + \rho_\theta(m\Delta\nu)) \;=\; \frac{\rho_\theta(m\Delta\nu)}{m}. \tag{3.1}$$

This equation provides support points for standard interpolation or extrapolation formulae. For example, if we know $\rho_\theta(\pm m\Delta\nu)$ for a given m, simple linear interpolation for σ yields:

$$\sigma(\theta, \eta) \;\approx\; \frac{1}{m}\left\{\frac{\rho_\theta(m\Delta\nu) + \rho_\theta(-m\Delta\nu)}{\rho_\theta(m\Delta\nu) - \rho_\theta(-m\Delta\nu)}\;[\eta - \theta - \rho_\theta(-m\Delta\nu)] \;-\; \rho_\theta(-m\Delta\nu)\right\}. \tag{3.2}$$

In the following we will refer to this equation as a 1^{st} order formula for σ. A 0^{th} order formula can be obtained by approximating $\sigma(\theta, \eta)$ by its value at $\eta = \theta$:

$$\sigma(\theta, \eta) \;\approx\; \sigma(\theta, \theta) \;=\; \rho'_\theta(0)\,\Delta\nu \;\approx\; \rho_\theta(\Delta\nu), \tag{3.3}$$

where $\rho'_\theta(\epsilon)$ is the derivative of $\rho_\theta(\epsilon)$ with respect to ϵ, and the last approximation on the right-hand side is valid for small $\Delta\nu$.

The Jacobian in the integrand of equation (2.25) is more difficult to approximate due to the presence of a derivative. Using equation (2.24), this Jacobian can be rewritten as:

$$\mathcal{J}(\theta,\eta) \;\equiv\; 1 \,+\, \frac{\eta-\theta}{\sigma(\theta,\eta)}\,\frac{\partial\sigma(\theta,\eta)}{\partial\theta}, \tag{3.4}$$

$$\;=\; \frac{\sigma(\theta,\eta)}{\Delta\nu}\;\left.\frac{1+\partial\rho_\theta(\epsilon)/\partial\theta}{\partial\rho_\theta(\epsilon)/\partial\epsilon}\right|_{\epsilon=\rho_\theta^{-1}(\eta-\theta)}. \tag{3.5}$$

Therefore:

$$\mathcal{J}(\theta,\theta+\rho_\theta(m\Delta\nu)) \;=\; \frac{\rho_\theta(m\Delta\nu)}{m\Delta\nu\,\left[\partial\rho_\theta(\epsilon)/\partial\epsilon\right]_{\epsilon=m\Delta\nu}}\;\left[1\,+\,\frac{\partial\rho_\theta(m\Delta\nu)}{\partial\theta}\right]. \tag{3.6}$$

If it can be assumed that $\rho_\theta(\epsilon)$ is approximately linear around $\epsilon = m\Delta\nu$, then the first factor on the right-hand side is 1 and we obtain:

$$\mathcal{J}(\theta,\theta+\rho_\theta(m\Delta\nu)) \;\approx\; 1\,+\,\frac{\partial\rho_\theta(m\Delta\nu)}{\partial\theta}. \tag{3.7}$$

Together with the point $\mathcal{J}(\theta,\theta)=1$, linear interpolation gives then:

$$\mathcal{J}(\theta,\eta) \;\approx\; 1\,+\,\frac{\eta-\theta}{\rho_\theta(m\Delta\nu)}\,\frac{\partial\rho_\theta(m\Delta\nu)}{\partial\theta}. \tag{3.8}$$

Although it is difficult to determine how good the approximated Jacobian is in general, a partial correction is obtained by normalizing $p(\theta\,|\,x)$ to 1 over the physical region of θ, as would typically be done to extract upper limits for example.

3.2 Monte Carlo method

The Monte Carlo procedure for performing the integral (2.25) is as follows:

1. Draw a random number η from $\mathcal{L}^\star(\eta\,|\,\nu_0,x)$;
2. Draw a random number u from a Gaussian with mean 0 and width 1;
3. Repeat step 2 if $|u|>n$;
4. Solve the equation $\quad\theta=\eta+u\,\sigma(\theta,\eta)\quad$ for θ (see below);
5. Histogram θ.

Using the definition of $\sigma(\theta,\eta)$, the equation at step 4 can be rewritten in terms of the shift function, as follows:

$$\theta \;=\; \eta \;-\; \rho_\theta(-u\Delta\nu). \tag{3.9}$$

Suppose now that we know $\rho_\theta(m\Delta\nu)$ versus θ. Using the fact that $\rho_\theta(0)=0$ we can do a linear interpolation:

$$\rho_\theta(\epsilon) \;\approx\; \frac{\rho_\theta(m\Delta\nu)}{m\Delta\nu}\,\epsilon,$$

and the corresponding equation for θ is then:

$$\theta \;\approx\; \eta \;+\; \frac{u}{m}\,\rho_\theta(m\Delta\nu). \tag{3.10}$$

If we know both $\rho_\theta(m\Delta\nu)$ and $\rho_\theta(-m\Delta\nu)$ as functions of θ, then a quadratic interpolation yields:

$$\rho_\theta(\epsilon) \;\approx\; \frac{\rho_\theta(m\Delta\nu)+\rho_\theta(-m\Delta\nu)}{2}\Big[\frac{\epsilon}{m\Delta\nu}\Big]^2 \;+\; \frac{\rho_\theta(m\Delta\nu)-\rho_\theta(-m\Delta\nu)}{2}\Big[\frac{\epsilon}{m\Delta\nu}\Big],$$

and the equation for θ becomes:

$$\theta \;\approx\; \eta + \frac{u}{m}\left(1-\frac{u}{m}\right)\frac{\rho_\theta(m\Delta\nu)}{2} \;-\; \frac{u}{m}\left(1+\frac{u}{m}\right)\frac{\rho_\theta(-m\Delta\nu)}{2}. \tag{3.11}$$

In contrast with quadrature methods, the Monte Carlo method only requires interpolation of one quantity, the shift $\rho_\theta(\epsilon)$ as a function of ϵ. As shown in Appendix A, the Jacobian is automatically taken into account by the Monte Carlo method, which is a significant advantage.

3.3 Generalization to more than one systematic uncertainty

To fix ideas we consider here a situation with three nuisance parameters, ν_1, ν_2, and ν_3. Inference about the interest parameter θ is then based on the multidimensional equivalent of equation (2.12):

$$p(\theta\,|\,x) \;\approx\; \int_{\nu_{1\circ}-n\Delta\nu_1}^{\nu_{1\circ}+n\Delta\nu_1} d\nu_1 \int_{\nu_{2\circ}-n\Delta\nu_2}^{\nu_{2\circ}+n\Delta\nu_2} d\nu_2 \int_{\nu_{3\circ}-n\Delta\nu_3}^{\nu_{3\circ}+n\Delta\nu_3} d\nu_3\; \mathcal{L}^\star(\theta\,|\,\nu_1,\nu_2,\nu_3,x) \\ \frac{e^{-\frac{1}{2}\left(\frac{\nu_3-\nu_{3\circ}}{\Delta\nu_3}\right)^2}}{\sqrt{2\pi}\,\Delta\nu_3}\;\frac{e^{-\frac{1}{2}\left(\frac{\nu_2-\nu_{2\circ}}{\Delta\nu_2}\right)^2}}{\sqrt{2\pi}\,\Delta\nu_2}\;\frac{e^{-\frac{1}{2}\left(\frac{\nu_1-\nu_{1\circ}}{\Delta\nu_1}\right)^2}}{\sqrt{2\pi}\,\Delta\nu_1}. \tag{3.12}$$

We now need three shift functions, $\rho_{1\theta}$, $\rho_{2\theta}$, and $\rho_{3\theta}$, one for each nuisance parameter:

$$\begin{aligned} P(\theta\,|\,\nu_{1\circ}+\epsilon_1,\nu_{2\circ},\nu_{3\circ},x) &= P(\theta+\rho_{1\theta}(\epsilon_1)\,|\,\nu_{1\circ},\nu_{2\circ},\nu_{3\circ},x),\\ P(\theta\,|\,\nu_{1\circ}+\epsilon_1,\nu_{2\circ}+\epsilon_2,\nu_{3\circ},x) &= P(\theta+\rho_{2\theta}(\epsilon_1,\epsilon_2)\,|\,\nu_{1\circ}+\epsilon_1,\nu_{2\circ},\nu_{3\circ},x),\\ P(\theta\,|\,\nu_{1\circ}+\epsilon_1,\nu_{2\circ}+\epsilon_2,\nu_{3\circ}+\epsilon_3,x) &= P(\theta+\rho_{3\theta}(\epsilon_1,\epsilon_2,\epsilon_3)\,|\,\nu_{1\circ}+\epsilon_1,\nu_{2\circ}+\epsilon_2,\nu_{3\circ},x), \end{aligned}$$

where P is, as before, the cumulative integral of $\mathcal{L}^\star(\theta\,|\,\nu_1,\nu_2,\nu_3,x)$ with respect to θ. Note that the above definitions assign a different number of arguments to each shift function. However, we shall assume that the shift function corresponding to a given nuisance parameter does not depend on the values of the other nuisance parameters, i.e. that $\rho_{2\theta}(\epsilon_1,\epsilon_2)$ does not depend on ϵ_1 and that $\rho_{3\theta}(\epsilon_1,\epsilon_2,\epsilon_3)$ does not depend on ϵ_1 or ϵ_2. The notation will be simplified accordingly:

$$\rho_{2\theta}(\epsilon_1,\epsilon_2) \;\equiv\; \rho_{2\theta}(\epsilon_2), \tag{3.13a}$$

$$\rho_{3\theta}(\epsilon_1,\epsilon_2,\epsilon_3) \;\equiv\; \rho_{3\theta}(\epsilon_3). \tag{3.13b}$$

Next, we apply the procedure described in section 2, from equation (2.16) to equation (2.25), to each integral on the right-hand side of (3.12), starting with the inner integral over ν_3. The end-result of this calculation is:

$$p(\theta \,|\, x) \;\approx\; \int_{\theta+\rho_{3\theta}(-n\Delta\nu_3)}^{\theta+\rho_{3\theta}(n\Delta\nu_3)} d\eta_3 \int_{\eta_3+\rho_{2\eta_3}(-n\Delta\nu_2)}^{\eta_3+\rho_{2\eta_3}(n\Delta\nu_2)} d\eta_2 \int_{\eta_2+\rho_{1\eta_2}(-n\Delta\nu_1)}^{\eta_2+\rho_{1\eta_2}(n\Delta\nu_1)} d\eta_1 \; \mathcal{L}^{\star}(\eta_1 \,|\, \nu_{1\circ}, \nu_{2\circ}, \nu_{3\circ}, x)$$

$$\cdot \frac{e^{-\frac{1}{2}\left[\frac{\eta_1-\eta_2}{\sigma_1(\eta_2,\eta_1)}\right]^2}}{\sqrt{2\pi}\,\sigma_1(\eta_2,\eta_1)} \; \frac{e^{-\frac{1}{2}\left[\frac{\eta_2-\eta_3}{\sigma_2(\eta_3,\eta_2)}\right]^2}}{\sqrt{2\pi}\,\sigma_2(\eta_3,\eta_2)} \; \frac{e^{-\frac{1}{2}\left[\frac{\eta_3-\theta}{\sigma_3(\theta,\eta_3)}\right]^2}}{\sqrt{2\pi}\,\sigma_3(\theta,\eta_3)}$$

$$\cdot \left[1 + \frac{\eta_1-\eta_2}{\sigma_1(\eta_2,\eta_1)} \frac{\partial\sigma_1(\eta_2,\eta_1)}{\partial\eta_2}\right] \left[1 + \frac{\eta_2-\eta_3}{\sigma_2(\eta_3,\eta_2)} \frac{\partial\sigma_2(\eta_3,\eta_2)}{\partial\eta_3}\right] \left[1 + \frac{\eta_3-\theta}{\sigma_3(\theta,\eta_3)} \frac{\partial\sigma_3(\theta,\eta_3)}{\partial\theta}\right], \tag{3.14}$$

where:

$$\sigma_i(\theta,\eta) \;\equiv\; \frac{\eta-\theta}{\rho_{i\theta}^{-1}(\eta-\theta)}\,\Delta\nu_i, \quad i = 1,2,3. \tag{3.15}$$

Equation (3.14) is the three-dimensional equivalent of (2.25). With the assumption of independence expressed by equations (3.13), the procedure for incorporating several systematics into a likelihood function is straightforward. First, calculate the shift function for each nuisance parameter separately, keeping the other nuisance parameters at their nominal values. The convolutions can then be performed in succession, the output of one convolution being used as input for the next. For the Monte Carlo method described in section 3.2 and assuming a total of s systematic uncertainties, this gives the following general algorithm:

1. Draw a random number η_1 from $\mathcal{L}^{\star}(\eta_1 \,|\, \nu_0, x)$;
2. Set $i = 1$;
3. Draw a random number u_i from a Gaussian with mean 0 and width 1;
4. Repeat the previous step if $|u_i| > n$;
5. Solve $\eta_{i+1} = \eta_i + u_i\,\sigma_i(\eta_{i+1}, \eta_i)$ for η_{i+1};
6. Increment i;
7. If $i \leq s$, go back to step 3;
8. Set $\theta = \eta_{s+1}$ and histogram θ.

The equation at step 5 can of course be handled in the same way as in problems with only one systematic uncertainty.

4 Example: a Truncated Gaussian Likelihood

To illustrate the method, we consider a simple problem with one interest parameter θ, one nuisance parameter ν, and a physical constraint $\theta \geq 0$. The probability density function of the measurement x is Gaussian:

$$f(x \,|\, \theta, \nu) \;=\; \frac{e^{-\frac{1}{2}\left(\frac{\theta-\nu-x}{\tau}\right)^2}}{\sqrt{2\pi}\,\tau}. \tag{4.1}$$

The corresponding normalized likelihood function is a truncated Gaussian density:

$$\mathcal{L}^{\star}(\theta \,|\, \nu, x) \;=\; \frac{e^{-\frac{1}{2}\left(\frac{\theta-\nu-x}{\tau}\right)^2}}{\sqrt{2\pi}\,\tau\,\frac{1}{2}\left[1+\mathrm{erf}\left(\frac{\nu+x}{\sqrt{2}\,\tau}\right)\right]} \qquad (\theta \geq 0), \tag{4.2}$$

where the normalization is done with the help of the error function:

$$\mathrm{erf}(z) \;\equiv\; \frac{2}{\sqrt{\pi}} \int_0^z dt\; e^{-t^2}. \tag{4.3}$$

The corresponding cumulative distribution is:

$$P(\theta \,|\, \nu) \;=\; \frac{\mathrm{erf}\left(\frac{\theta-\nu-x}{\sqrt{2}\,\tau}\right) \;+\; \mathrm{erf}\left(\frac{\nu+x}{\sqrt{2}\,\tau}\right)}{1 \;+\; \mathrm{erf}\left(\frac{\nu+x}{\sqrt{2}\,\tau}\right)}. \tag{4.4}$$

Next, we assume that prior information about ν comes from a subsidiary Gaussian measurement $\nu_0 \pm \Delta\nu$:

$$\pi(\nu) \;=\; \frac{e^{-\frac{1}{2}\left(\frac{\nu-\nu_0}{\Delta\nu}\right)^2}}{\sqrt{2\pi}\,\Delta\nu}, \tag{4.5}$$

and that there is no prior information about θ, so that a uniform prior is chosen: $\pi(\theta) = 1$. With these ingredients the predictive density, equation (2.2), is given by:

$$m(x) \;=\; \frac{1}{2}\left[1 \;+\; \mathrm{erf}\left(\frac{\nu_0 + x}{\sqrt{2}\,\sqrt{\Delta\nu^2+\tau^2}}\right)\right], \tag{4.6}$$

and the marginal posteriors for θ and ν, equations (2.3) and (2.4) respectively, are:

$$p(\theta \,|\, x) \;=\; \frac{e^{-\frac{1}{2}\left(\frac{\theta-\nu_0-x}{\sqrt{\Delta\nu^2+\tau^2}}\right)^2}}{\sqrt{2\pi\,(\Delta\nu^2+\tau^2)}\;\frac{1}{2}\left[1 \;+\; \mathrm{erf}\left(\frac{\nu_0+x}{\sqrt{2\,(\Delta\nu^2+\tau^2)}}\right)\right]} \tag{4.7}$$

$$p(\nu \,|\, x) \;=\; \frac{1+\mathrm{erf}\left(\frac{\nu+x}{\sqrt{2}\,\tau}\right)}{1+\mathrm{erf}\left(\frac{\nu_0+x}{\sqrt{2\,(\Delta\nu^2+\tau^2)}}\right)}\;\frac{e^{-\frac{1}{2}\left(\frac{\nu-\nu_0}{\Delta\nu}\right)^2}}{\sqrt{2\pi}\,\Delta\nu}. \tag{4.8}$$

These are exact formulae for the example considered. Before testing the approximation of the convolution method, we note that the nuisance posterior $p(\nu \,|\, x)$ equals the

nuisance prior $\pi(\nu)$ in the limit of large $\nu_0 + x$. This can in fact be interpreted as a consistency condition on the priors and the data. Indeed, the likelihood favors values of the data x that are close to $\theta - \nu$, and the nuisance prior favors values of the subsidiary measurement ν_0 that are close to ν. Hence we have $x \sim \theta - \nu_0$, so that the physical constraint $\theta \geq 0$ implies $x + \nu_0 \geq 0$. Given that x and ν_0 are measurement results that fluctuate with standard deviations τ and $\Delta\nu$ respectively, actual consistency between data and priors requires $\nu_0 + x$ to be large relatively to $\sqrt{\Delta\nu^2 + \tau^2}$. For the remainder of this section we will choose $x = 0.0$, $\tau = 0.5$, $\nu_0 = 1.0$, and $\Delta\nu = 0.1$, 0.2. Figure 3 shows that with this choice there is hardly any difference between the nuisance prior and posterior. Hence, condition (2.9) is satisfied and the convolution method can be applied.

The normalized likelihood function and corresponding cumulative distribution are shown in Figures 4 and 5 respectively, as functions of θ, for $\nu = 0.8$, 1.0, and 1.2.

The shift function $\rho_\theta(\epsilon)$ is shown as a function of θ in Figure 6. This curve becomes flat as one moves away from the physical boundary, in agreement with the behavior of the shift function in the corresponding situation without a physical boundary (see Figure 2, top). Since the physical boundary does not depend on the nuisance parameter, the shift function goes to zero as θ approaches zero. The ϵ dependence of $\rho_\theta(\epsilon)$ is examined in Figure 7. As expected from definition (2.13), $\rho_\theta(\epsilon)$ goes through the point $(\epsilon = 0, \rho_\theta = 0)$ and is approximately linear in the vicinity of that point. The flat portion of each curve is the effect of the physical boundary: one has $\lim_{\epsilon\to\infty} \rho_\theta(\epsilon) = -\theta$. This limit is reached in such a way that for each θ, $\rho_\theta(\epsilon)$ is a one-to-one function of ϵ from $]-\infty, +\infty[$ to $]-\theta, +\infty[$, as required for the change of variable (2.18) to be valid.

The exact width $\sigma(\theta, \eta)$ of the convolution kernel, equation (2.21), is shown in Figure 8 as a function of η. Superimposed on that plot are horizontal line segments indicating, for a given θ value, the region of integration $[\theta + \rho_\theta(-n\Delta\nu), \theta + \rho_\theta(n\Delta\nu)]$, with $n = 3$ and $\Delta\nu = 0.1$, 0.2. The plot shows that the width is approximately constant over the region of integration, and more so at small $\Delta\nu$ and large θ.

Figure 9 shows the correct Jacobian factor derived from the $\sigma(\theta, \eta)$ of the previous plot, again with the integration regions indicated by horizontal line segments. Each Jacobian equals 1 near the middle of the corresponding integration region and remains fairly constant over the whole region, especially at high θ and small $\Delta\nu$.

Figure 10 shows the final result of using a quadrature method to fold a systematic uncertainty $\Delta\nu = 0.2$ into the likelihood of equation (4.2). The top plot compares the unsmeared likelihood with the smeared one, using the "exact" formula, equation (4.7). The bottom plot compares the "exact" smeared likelihood with convolution approximations based on 0^{th} order (dashed curve) and 1^{st} order (with $m = 1$, dotted curve) formulae for the width σ of the convolution kernel. The integration range of the convolution corresponds to $n = 3$. The approximations are generally good, except at low θ. This is expected since the width and Jacobian factors are also less well modeled by the approximations at low θ. Note however that the 1^{st} order formula for σ performs better than the 0^{th} order one. Lastly, Figure 11 shows the result of applying the Monte Carlo method to the same problem. The exact smeared likelihood is compared with Monte Carlo curves (each obtained from 10^8 runs of the algorithm) based on 1^{st} and 2^{nd} order interpolations for the shift function (equations (3.10) and (3.11) respectively).

We emphasize that in practical applications of the method, Figures 7, 8, and 9 are *not* plotted. In fact, in most cases these plots would be very difficult to produce since they require exact knowledge of the likelihood function dependence on the nuisance parameter ν. They were shown in this example only to help explain the performance of the approximations used.

5 Conclusions

We have described a convolution method for incorporating systematic uncertainties into a likelihood function for the purpose of extracting upper limits or two-sided intervals. The method is very general in that it does not make any assumptions about the form of the likelihood function, other than that the latter be normalizable over the physical region of the parameter of interest. If for example the likelihood is derived from a fit to some spectrum, then the method works equally well whether the fit is binned or unbinned. An alternative method for incorporating systematic uncertainties in the case of binned fits is described in Reference [5].

The main advantage of the method proposed here is that it is derived from an exact Bayesian formulation of the problem. As a result, there is no arbitrariness in the expression for the width σ of the convolution kernel, equation (2.21). This is in marked contrast with other approaches to this problem, which rely on intuition and guesswork to fix the form of σ.

The core of the method is the determination of the shift function, $\rho_\theta(m\Delta\nu)$, for one or more values of m. In principle it should be possible to improve the approximations by increasing the number of values of m for which the shift function is available. In that case however, one needs to make a careful choice of interpolation method, as high-order polynomial interpolation may exhibit more variation than desired. On the other hand this is not a real problem in most applications, where it is computationally too expensive to determine the shape of the likelihood at more than a few values of the nuisance parameter(s).

Acknowledgements

I wish to thank Hyunsoo Kim and Jieun Kim for inducing me to write this note and for providing feedback based on their attempt to use this method.

Appendix

A Justification of the Monte Carlo algorithm

The notation in this appendix refers to the algorithm description at the beginning of section 3.2. Since η and u are independent random numbers, their joint density factorizes:

$$f(\eta, u) \;=\; \mathcal{L}^{\star}(\eta \,|\, \nu_0, x)\, \frac{e^{-u^2/2}}{\sqrt{2\pi}\, K}, \tag{A.1}$$

where $K \equiv \mathrm{erf}(n/\sqrt{2})$ corrects the normalization for the fact that u is required to be between $-n$ and $+n$. The cumulative probability for the random number θ to be less than some constant θ_0 is then given by:

$$\Pr(\theta \leq \theta_0) \;=\; \iint_{\mathcal{R}} d\eta\, du\, \mathcal{L}^{\star}(\eta \,|\, \nu_0, x)\, \frac{e^{-u^2/2}}{\sqrt{2\pi}\, K}, \tag{A.2}$$

where the integration region $\mathcal{R}$ is defined by:

$$\begin{aligned} -n \;&\leq\; u \;\leq\; +n, \\ \theta(\eta, u) \;&\leq\; \theta_0, \end{aligned}$$

and the function $\theta(\eta, u)$ is implicitly defined by the constraint: $\eta + u\,\sigma(\theta, \eta) = \theta$. Next, we change integration variables:

$$u \to t: \quad u \;=\; \frac{t - \eta}{\sigma(t, \eta)}. \tag{A.3}$$

The corresponding transformation of differentials is:

$$du \;=\; \frac{dt}{\sigma(t, \eta)} \;-\; \frac{t - \eta}{\sigma(t, \eta)^2}\, \frac{\partial \sigma}{\partial t}\, dt, \tag{A.4}$$

and the integration region $\mathcal{R}$ transforms into:

$$\begin{aligned} t \;&\leq\; \theta_0, \\ t + \rho_t(-n\Delta\nu) \;&\leq\; \eta \;\leq\; t + \rho_t(n\Delta\nu), \end{aligned}$$

where we assumed that $\rho_t(\epsilon)$ increases with ϵ. Equation (A.2) becomes now:

$$\Pr(\theta \leq \theta_0) \;=\; \int_{-\infty}^{\theta_0} dt \int_{t+\rho_t(n\Delta\nu)}^{t+\rho_t(n\Delta\nu)} d\eta\, \mathcal{L}^{\star}(\eta \,|\, \nu_0, x)\, \frac{e^{-\frac{1}{2}\left(\frac{t-\eta}{\sigma(t,\eta)}\right)^2}}{\sqrt{2\pi}\, \sigma(t, \eta)\, K} \left[1 \;+\; \frac{\eta - t}{\sigma(t, \eta)}\, \frac{\partial \sigma(t, \eta)}{\partial t}\right]. \tag{A.5}$$

Differentiating this expression with respect to θ_0 yields equation (2.25), as required.

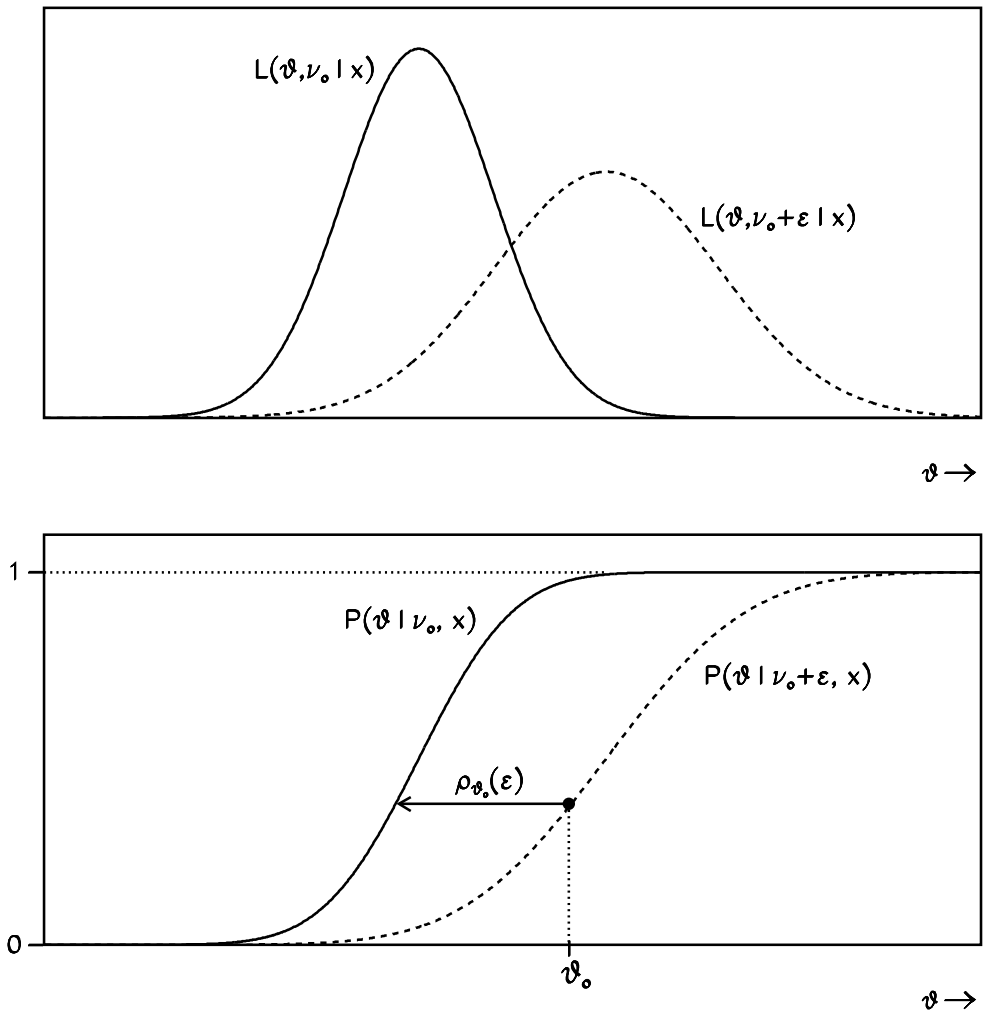

Figure 1: Illustration of the graphical construction method of the shift function $\rho_\theta(\epsilon)$. Top: unshifted (solid) and shifted (dashes) likelihood functions. Bottom: unshifted and shifted cumulative distributions derived from the top plot. For a given $\theta_\circ$ and ϵ, $\rho_{\theta_\circ}(\epsilon)$ is constructed as the signed horizontal distance between the shifted and unshifted cumulative distributions.

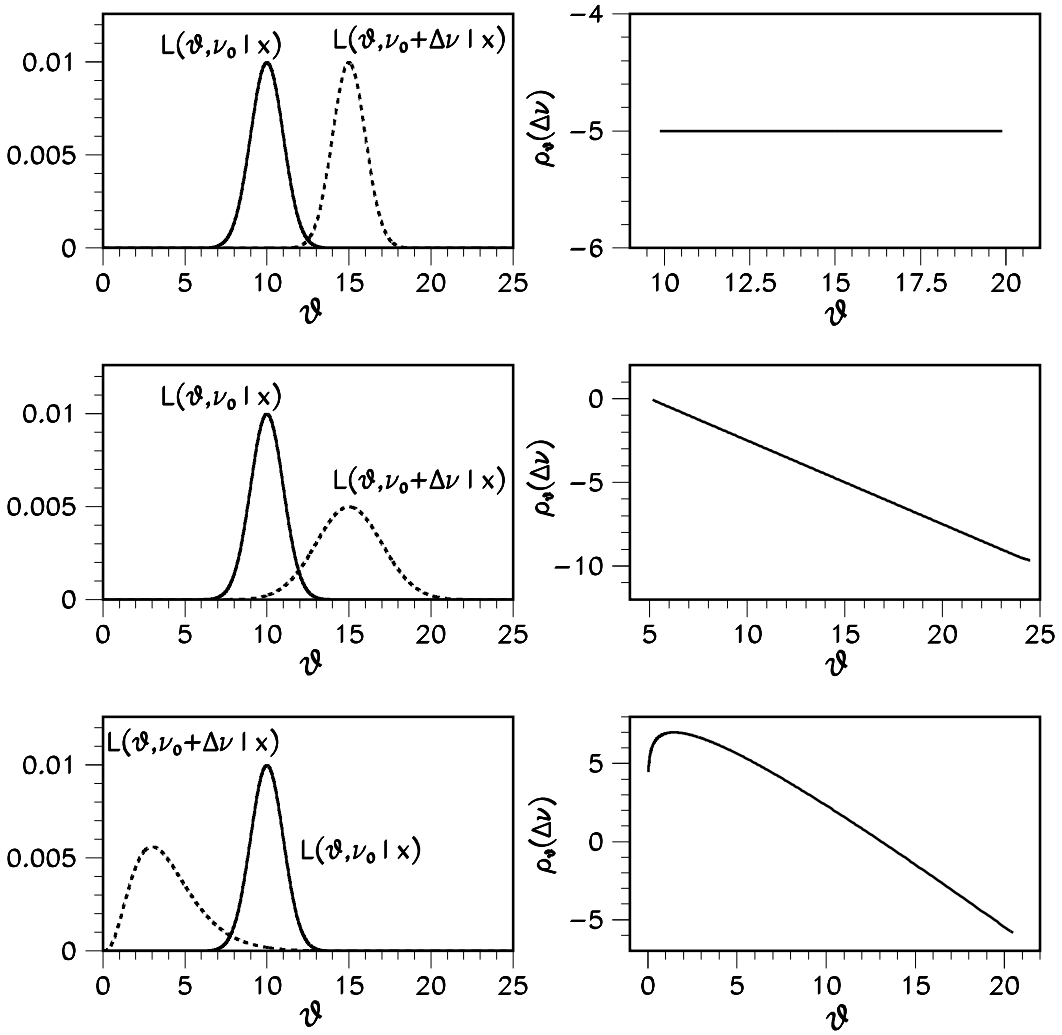

Figure 2: Examples of shift functions are given in the three right-hand side plots. They correspond to the unshifted and shifted likelihoods drawn on the left-hand side with solid and dashed lines respectively. Three cases are shown: unshifted and shifted likelihoods are Gaussians with same width but different mean (top); Gaussians with different width and mean (center); Gaussian and gamma (bottom).

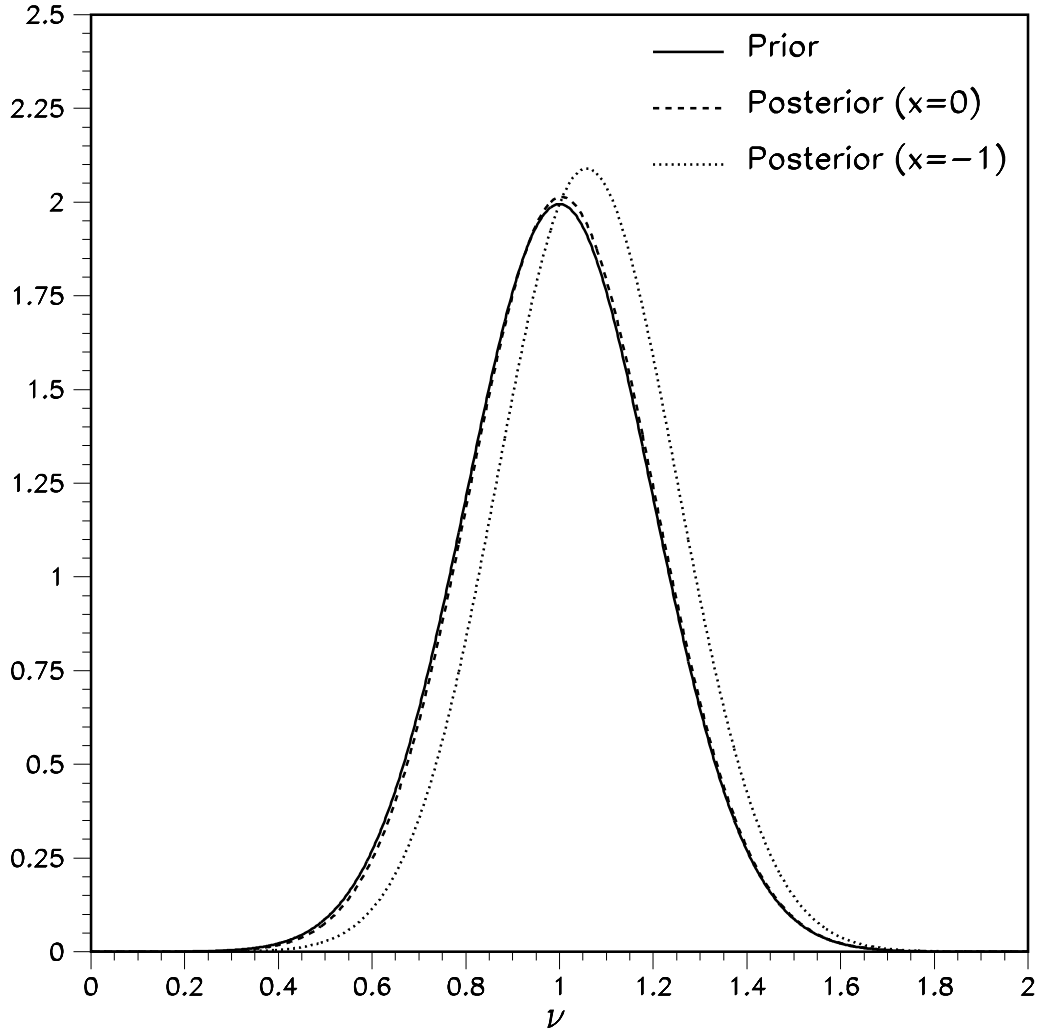

Figure 3: Marginal nuisance prior for $\nu_0 = 1.0$ and $\Delta\nu = 0.2$ (solid line), and corresponding marginal nuisance posteriors for $x = 0$ (dashed line) and $x = -1$ (dotted line). Only the value $x = 0$ is used in the remaining plots of this note. The curve with $x = -1$ illustrates the effect of a disagreement between likelihood and priors, in which case the convolution method may not be a good approximation.

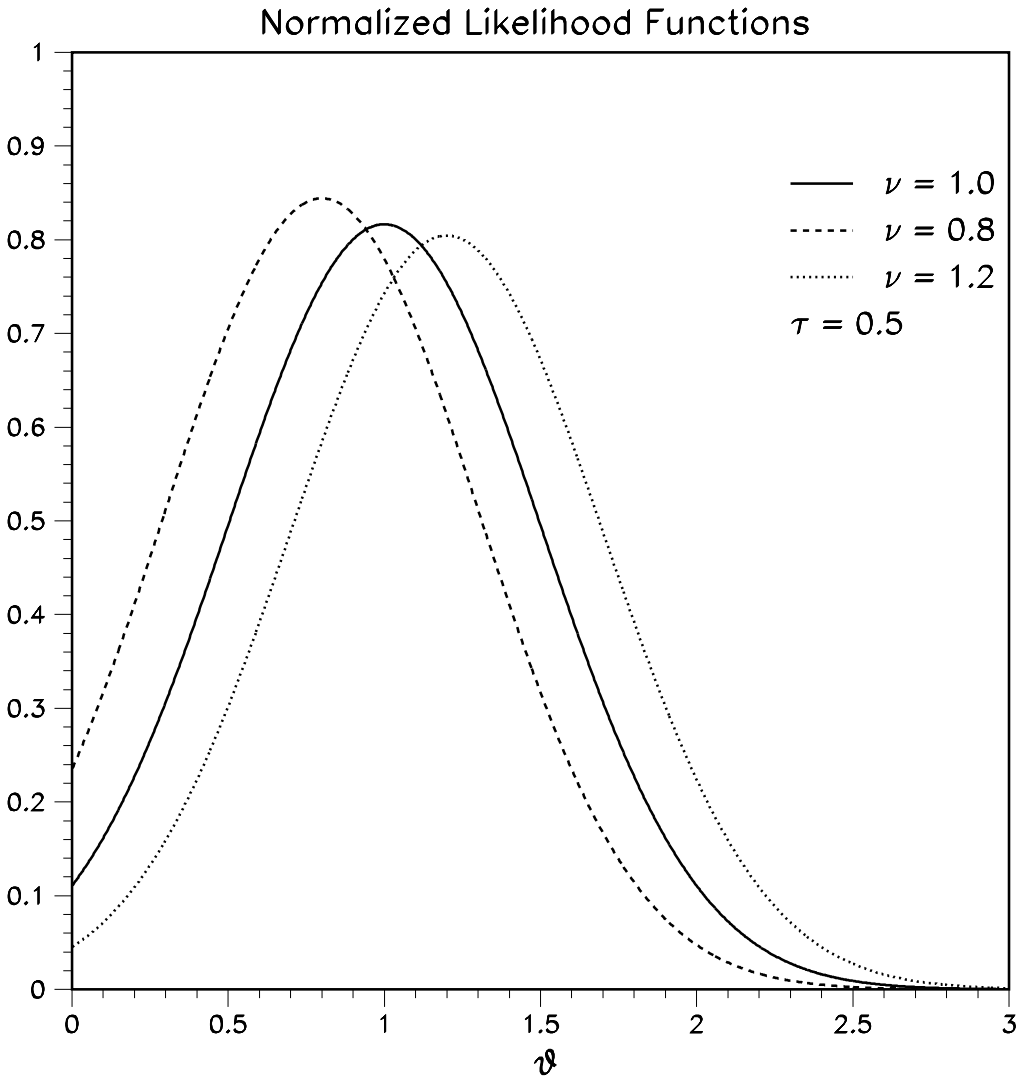

Figure 4: Shifted and unshifted truncated Gaussian likelihood functions, from the example discussed in section 4 of the text. The three curves are normalized to 1 over the physical region ($\theta \geq 0$).

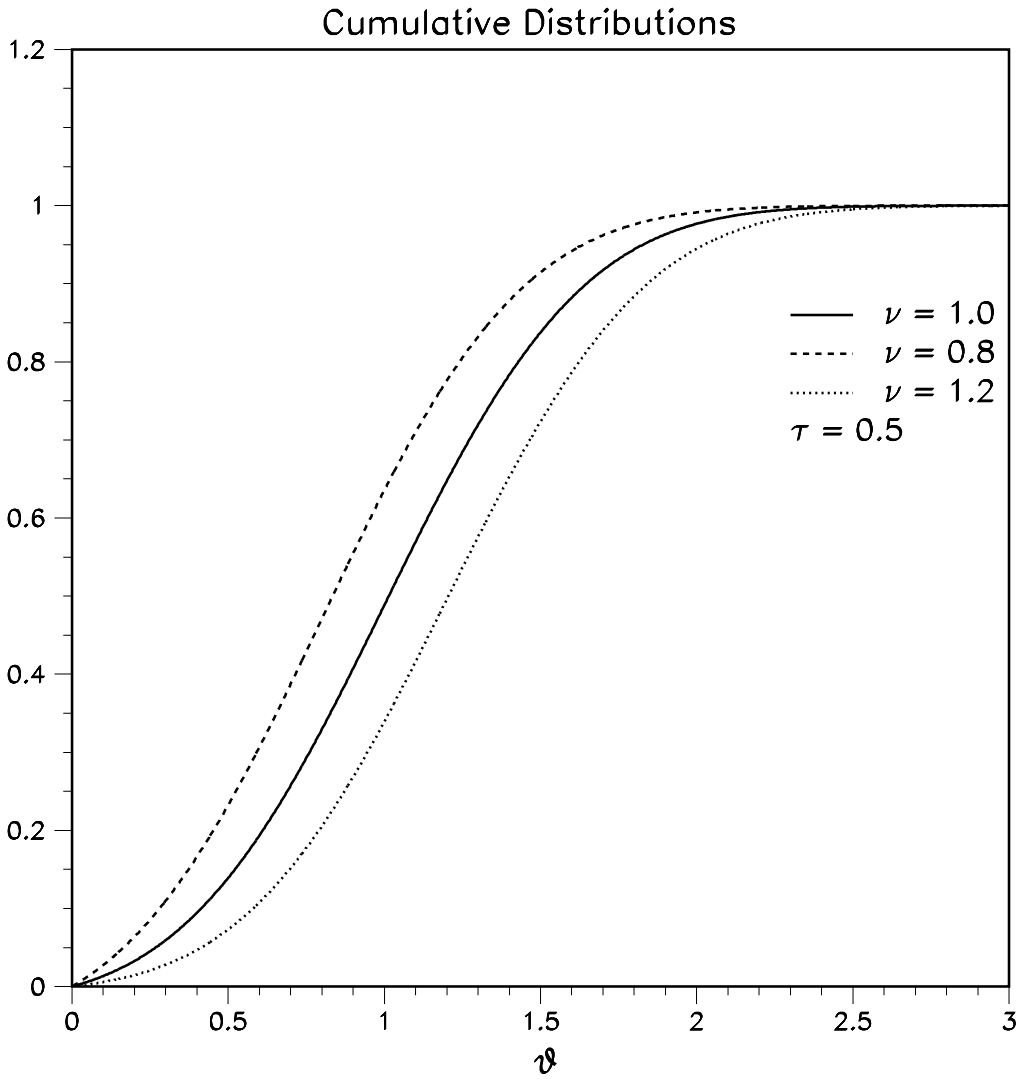

Figure 5: Shifted and unshifted cumulative distributions corresponding to the normalized likelihoods of Figure 4.

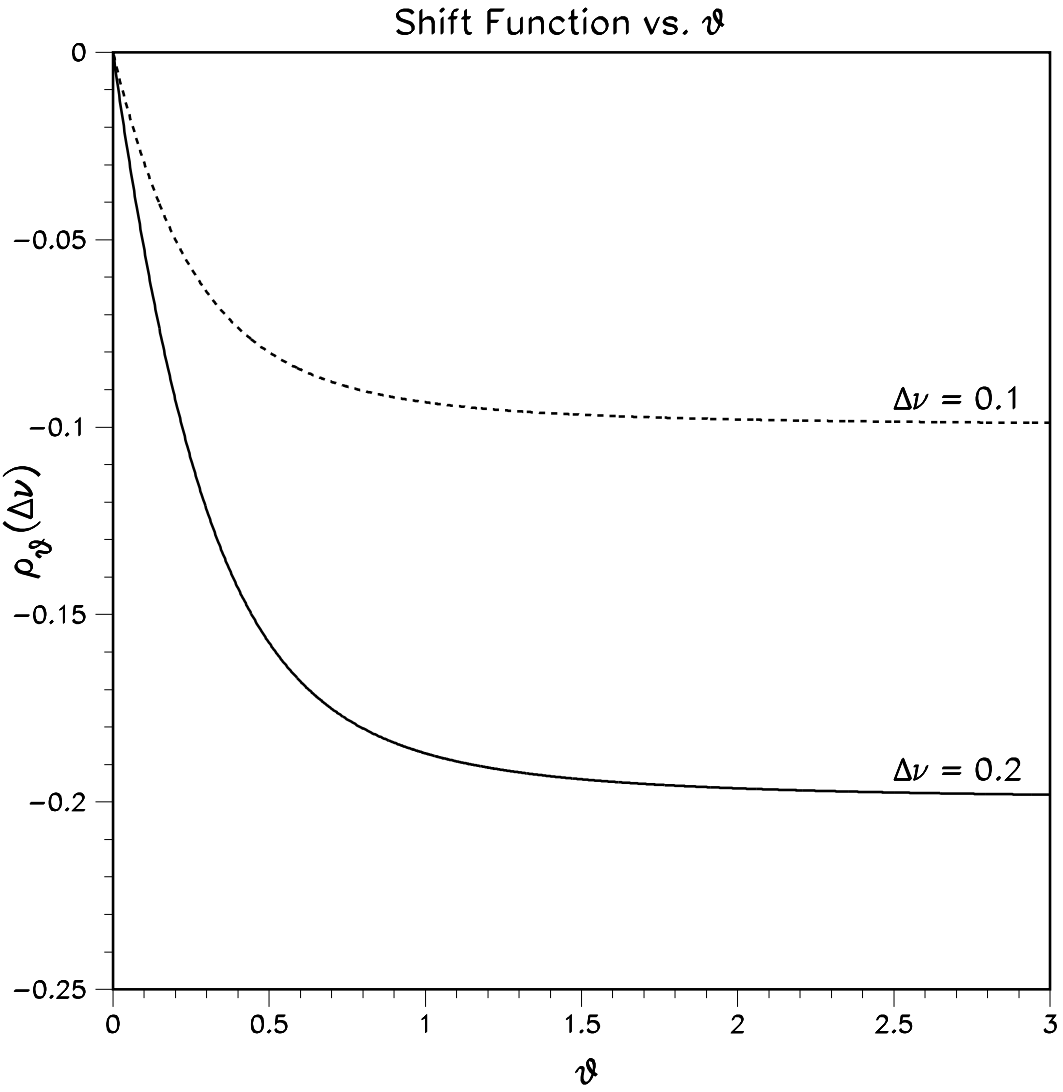

Figure 6: Shift function $\rho_\theta(\Delta\nu)$ versus θ, for $\Delta\nu = 0.1$ and 0.2. For $\Delta\nu = 0.2$ this function was determined from the $\nu = 1.2$ and $\nu = 1.0$ distributions shown in Figure 5.

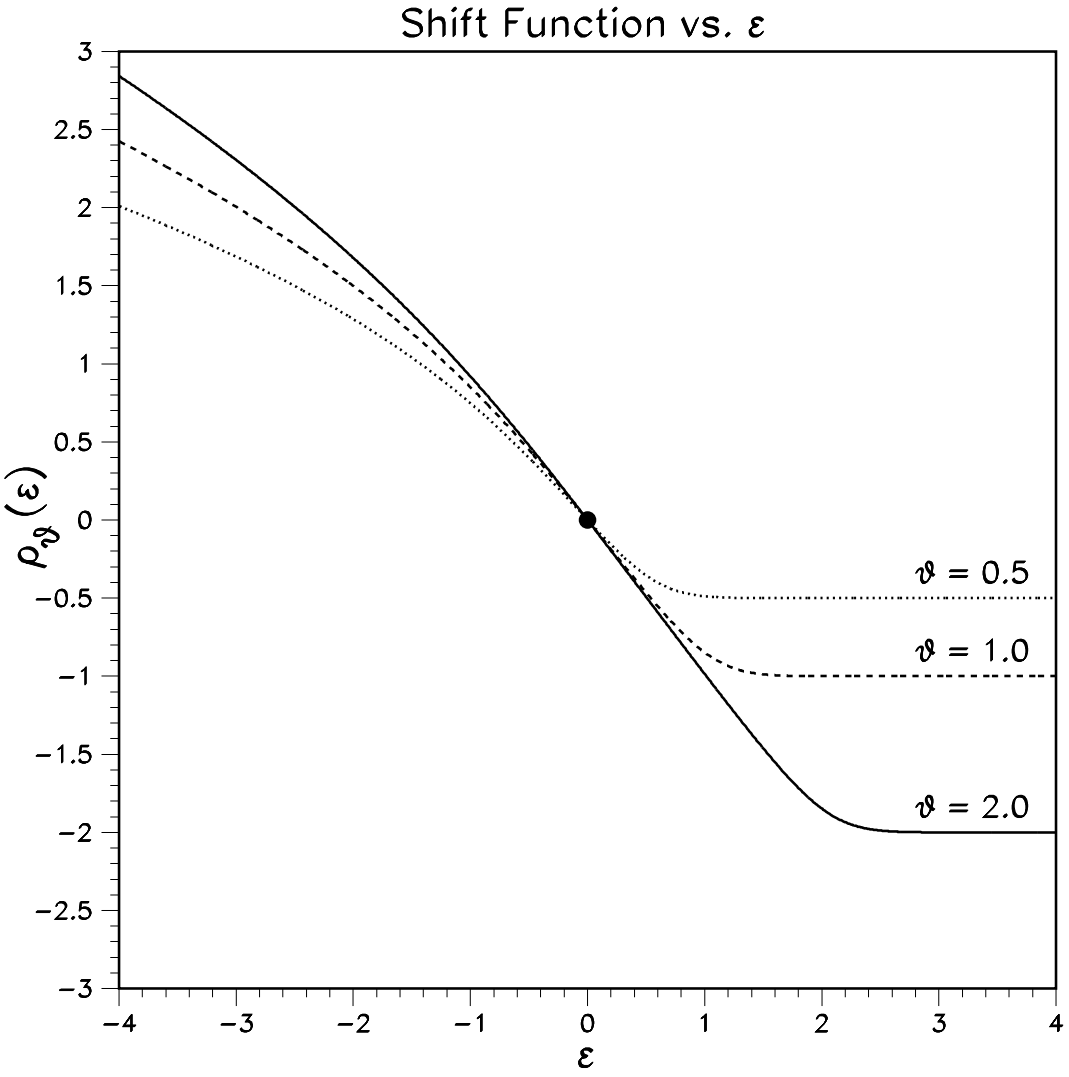

Figure 7: Shift functions $\rho_\theta(\epsilon)$ as a function of ϵ for three values of θ, for the example discussed in section 4. The large dot in the center is the origin of coordinates.

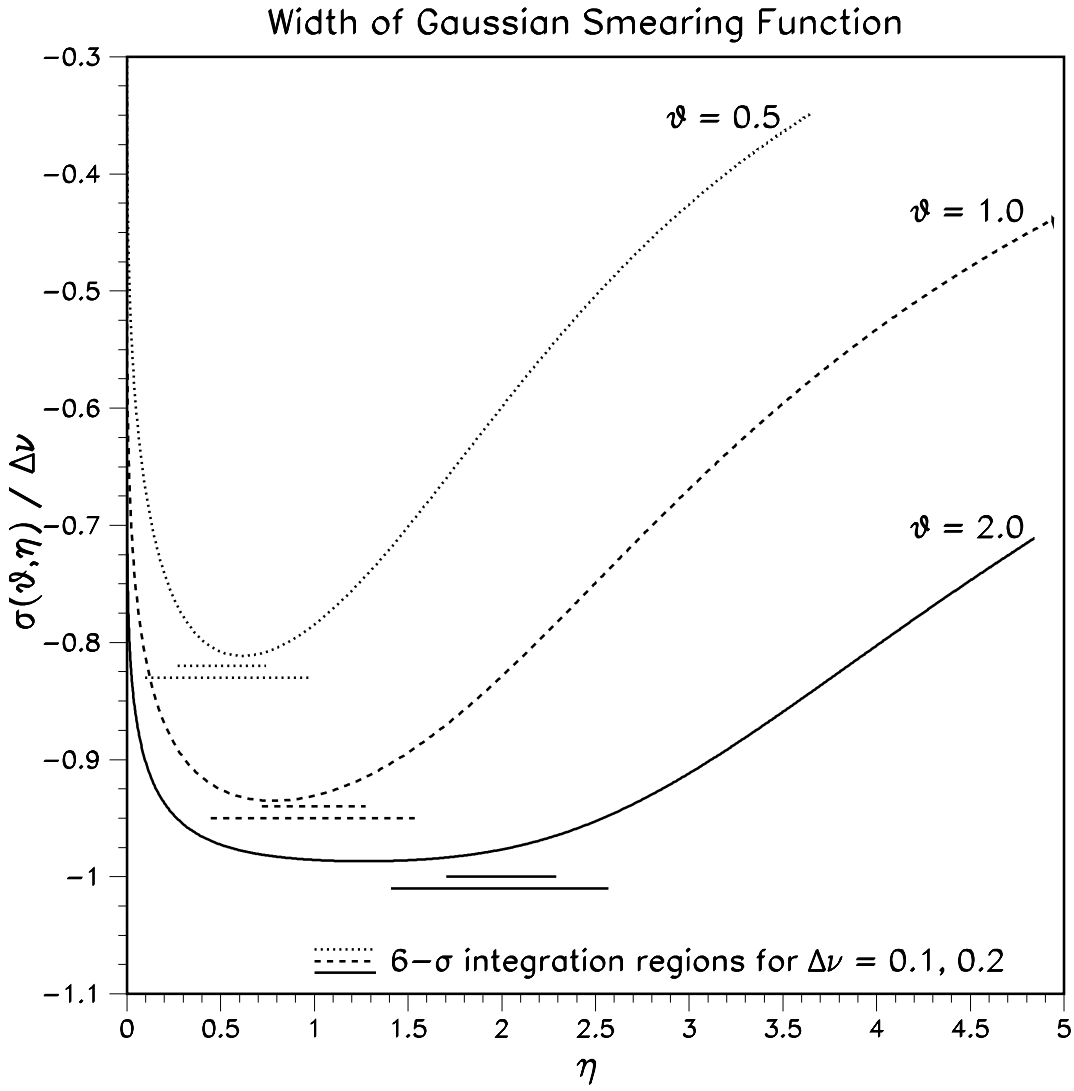

Figure 8: Scaled width of the Gaussian convolution kernel, $\sigma(\theta, \eta)/\Delta\nu$, as a function of η and for $\theta = 0.5$, 1.0, and 2.0. This is an exact calculation according to equation (2.21) in the text. The horizontal line segments indicate the integration region $[\theta + \rho_\theta(-n\Delta\nu), \theta + \rho_\theta(n\Delta\nu)]$ of equation (2.25), for $n = 3$ (hence the meaning of the notation "6σ" for the total length of the indicated regions).

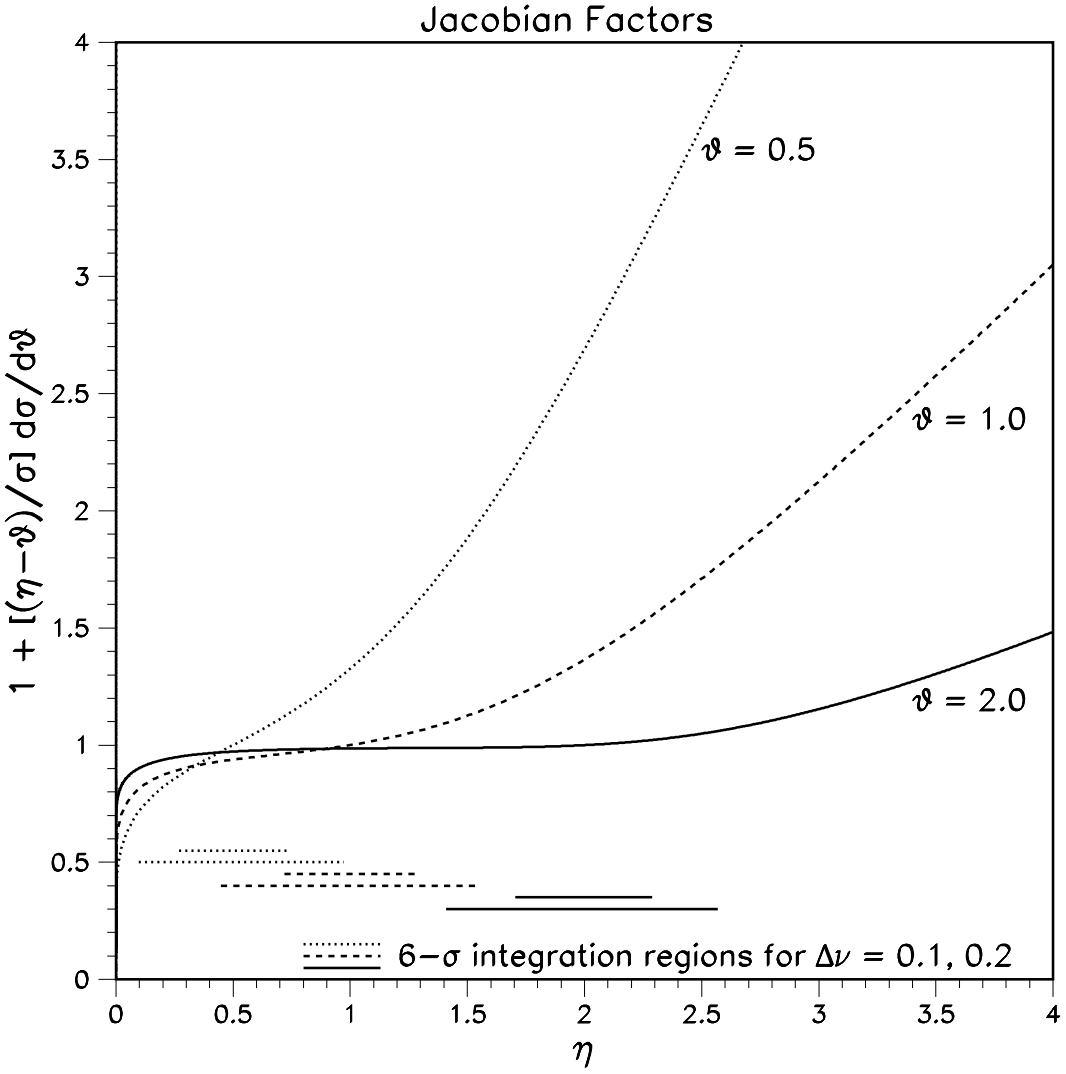

Figure 9: Exact Jacobian factor from the Gaussian convolution integral of equation (2.25). The horizontal line segments indicate the same integration regions as in Figure 8.

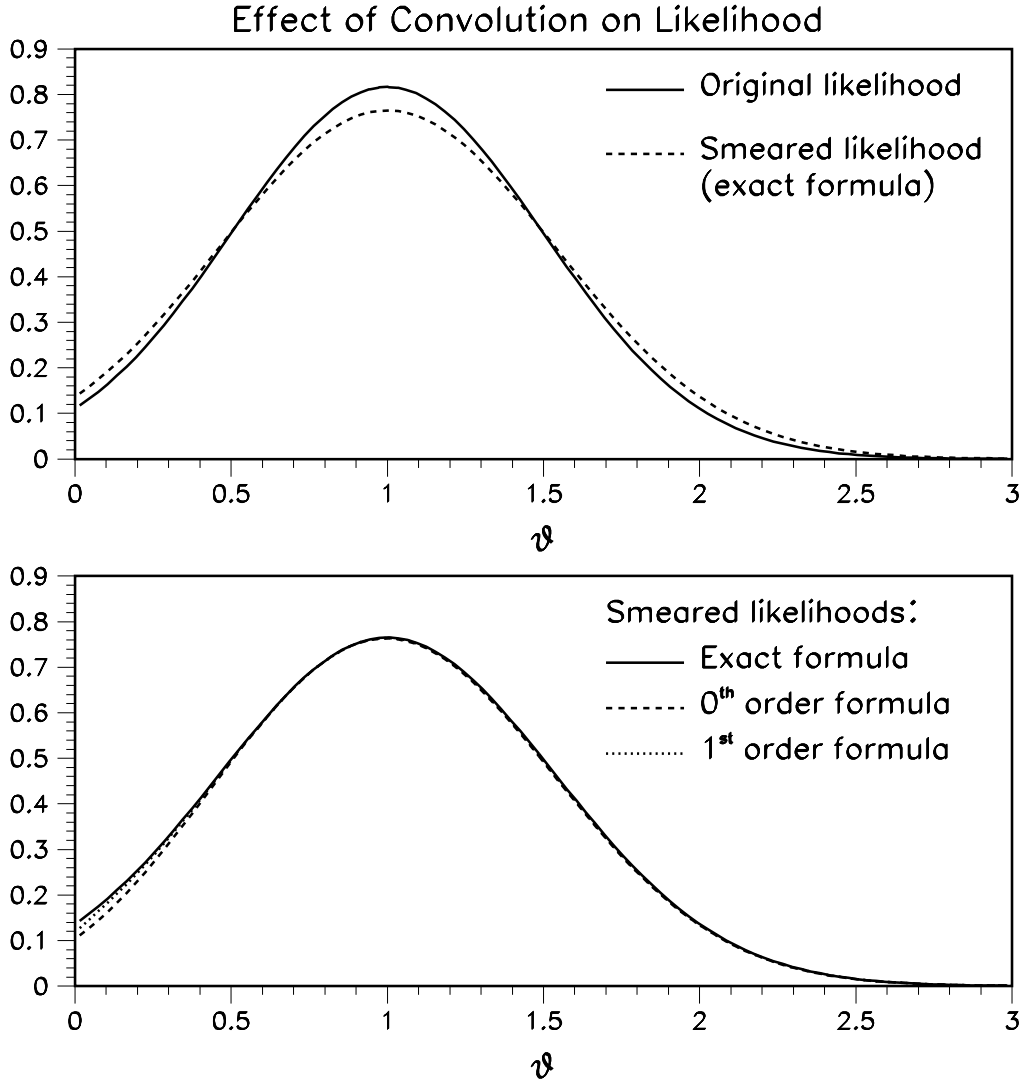

Figure 10: Effect of a $\Delta\nu = 0.2$ systematic uncertainty on the truncated Gaussian likelihood function of Figure 4 ($\nu_0 = 1.0$). Top: comparison of the unsmeared likelihood (solid line) with the likelihood smeared according to the exact formula (dashed line). Bottom: comparison of the likelihood smeared according to the exact formula (solid line) with quadrature approximations based on 0^{th} and 1^{st} order formulae for the width σ of the Gaussian convolution kernel (dashed and dotted lines, respectively). Note that these approximations are not normalized to 1 in order to facilitate comparison with the exact formula.

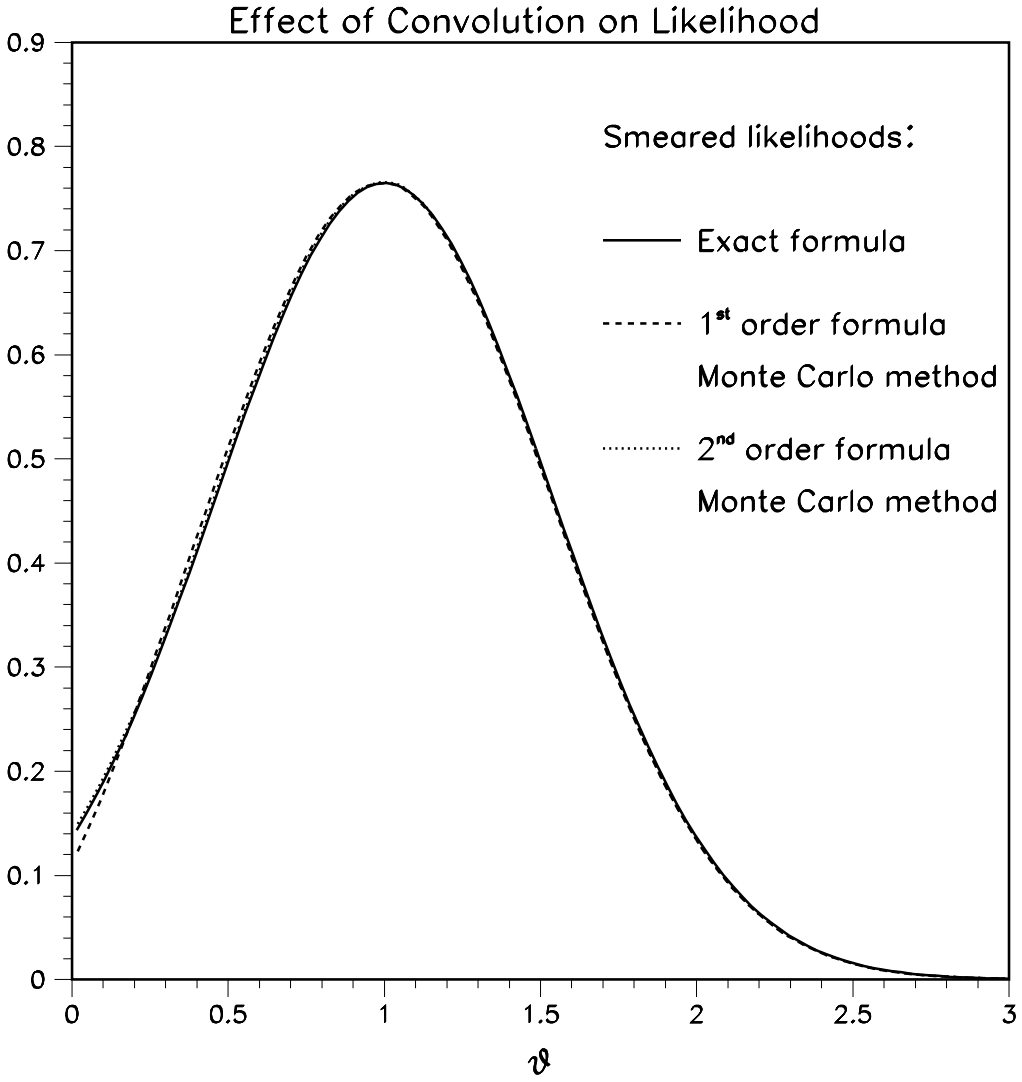

Figure 11: Effect of a $\Delta\nu = 0.2$ systematic uncertainty on the truncated Gaussian likelihood function of Figure 4 ($\nu_0 = 1.0$). The likelihood smeared according to the exact formula (solid line) is compared with Monte Carlo approximations based on 1^{st} and 2^{nd} order interpolation formulae for the shift function (dashed and dotted lines, respectively). All three curves have unit area. The two Monte Carlo curves were obtained from 10^8 runs of the algorithm in section 3.2.

References

[1] N. Reid, D.A.S. Fraser, "Likelihood inference in the presence of nuisance parameters," Proceedings of *PHYSTAT2003: Statistical Problems in Particle Physics, Astrophysics and Cosmology*, Stanford Linear Accelerator Center, Stanford, California, September 8-11, 2003; eConf C030908, SLAC-R-703. Also available as arXiv:physics/0312079 v1 (11 December 2003).

[2] James O. Berger, Brunero Liseo and Robert L. Wolpert, "Integrated likelihood methods for eliminating nuisance parameters," Statist. Sci. **14**, 1 (1999). Also available at `http://www.stat.duke.edu/~berger/papers/brunero.html`.

[3] Luc Demortier, "Bayesian treatments of systematic uncertainties," Proceedings of the Conference on *Advanced Statistical Techniques in Particle Physics*, Institute for Particle Physics Phenomenology, University of Durham, UK, 18-22 March 2002; IPPP/02/39 (2002), pg. 145. Also available at `http://www.ippp.dur.ac.uk/Workshops/02/statistics/proceedings/demortier.ps`.

[4] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling, "Numerical Recipes; The art of scientific computing," Cambridge University Press, 1986 (818pp).

[5] T. Devlin, "Correlations from systematic corrections to Poisson-distributed data in log-likelihood functions," CDF note 3126, version 3.1 (May 27, 1999).