BAYESIAN TREATMENTS OF SYSTEMATIC UNCERTAINTIES

Luc Demortier
The Rockefeller University, New York, NY 10021, U.S.A.

Abstract

The standard Bayesian treatment of systematic uncertainties is to integrate out the corresponding nuisance parameters from the joint posterior density for all parameters. We apply this formalism to measurements in which the data cannot distinguish between nuisance parameters and parameters of interest, and show that it leads to posterior densities with undesirable properties. To solve this problem, we propose to introduce correlations between the parameters of interest and the nuisance parameters *in the prior*, in such a way that the nuisance parameter information does not get updated by the measurement. Finally, we describe a method to replace Bayesian marginalization integrals over nuisance parameters by convolutions over the parameters of interest. Such convolutions are computationally more tractable and provide insight into some useful approximations.

1 INTRODUCTION

The Review of Particle Physics [1] describes the following method for incorporating systematic uncertainties in a Bayesian analysis. Suppose we have a likelihood $\mathcal{L}(x\,|\,\theta,\nu)$ expressing the probability density of the data x given a parameter of interest θ and a nuisance parameter ν. Because the data correlate ν and θ, a lack of knowledge about ν gives rise to a systematic uncertainty on θ. If $\pi(\theta,\nu)$ is the prior density, Bayes' theorem gives the posterior probability density as:

$$p(\theta,\nu\,|\,x) \;=\; \frac{1}{p(x)}\,\mathcal{L}(x\,|\,\theta,\nu)\,\pi(\theta,\nu), \tag{1}$$

where $p(x)$ is the marginal probability density of the data:

$$p(x) \;=\; \int d\nu \int d\theta\; \mathcal{L}(x\,|\,\theta,\nu)\,\pi(\theta,\nu). \tag{2}$$

To obtain a probability density for θ only, one simply integrates out the nuisance parameter ν from the joint posterior:

$$p(\theta\,|\,x) \;=\; \int d\nu\; p(\theta,\nu\,|\,x). \tag{3}$$

Systematic uncertainties are often such that the data alone cannot provide independent information about both θ and ν, as for example in the case of a cross section measurement in the presence of acceptance uncertainties. We discuss this example, and the difficulties it presents, in section 2. Section 3 offers a solution based on introducing correlations between parameters of interest and nuisance parameters in the prior. Finally, a method by which integrals over nuisance parameters are replaced by convolutions over parameters of interest is described in section 4.

2 CASE STUDY OF A CROSS SECTION MEASUREMENT WITH ACCEPTANCE UNCERTAINTIES

We apply the marginalization formalism to the measurement of a signal cross section σ in the presence of n observed events, b expected background events, an acceptance A with uncertainty ΔA, and an integrated luminosity L. The likelihood is given by the Poisson probability for observing n events:

$$\mathcal{L}(n\,|\,\sigma,A) \;=\; \frac{(\sigma LA+b)^n}{n!}\,e^{-\sigma LA-b}. \tag{4}$$

For obvious reasons we consider σ the parameter of interest and A the nuisance parameter. These parameters are generally treated as being *a priori* uncorrelated, so that their combined prior factorizes. We will further assume that the prior for A is a truncated Gaussian, whereas that for σ is flat:

$$\pi(\sigma, A) \;=\; \pi(\sigma)\,\pi(A) \;=\; H(\sigma)\,\frac{e^{-\frac{1}{2}\left(\frac{A-A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\,K\,\Delta A}\,H(A)\,H(1-A), \tag{5}$$

where H is Heaviside's step function ($H(x) = 1$ if $x \geq 0$, and 0 otherwise), and K is a normalization constant. Note that the above prior is improper (i.e. non-integrable) with respect to σ.

According to eq. (1), the calculation of posterior intervals or upper limits on σ requires a marginal data density that is finite. In the present case, we have from eq. (2):

$$p(n) \;=\; \int_0^1 dA \int_0^\infty d\sigma\;\mathcal{L}(n\,|\,\sigma, A)\;\pi(\sigma, A) \;=\; \frac{1}{L}\,e^{-b}\sum_{i=0}^{n}\frac{b^i}{i!}\int_0^1 dA\,\frac{1}{A}\frac{e^{-\frac{1}{2}\left(\frac{A-A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\,K\,\Delta A}. \tag{6}$$

Unfortunately, because of the $1/A$ factor in the integrand, this integral diverges at its lower limit (exchanging the order of integration does not remove the divergence). The posterior density is therefore improper and cannot be used to extract intervals [2]. An obvious solution is to regularize the cross section prior by introducing a cut-off $\sigma_{\max}$. The requirement $\sigma \leq \sigma_{\max}$ at the prior level then guarantees a proper posterior density, but posterior intervals and upper limits will depend on $\sigma_{\max}$. Regrettably, users of this method do not always quote their choice of $\sigma_{\max}$, making it difficult to interpret their results.

In the next subsections we consider alternative choices of prior to illustrate some issues and motivate a somewhat different treatment of systematic uncertainties in section 3.

2.1 Alternative prior 1: log-normal in A, flat in σ

Since the integrand of the marginal data density (6) has a non-integrable $1/A$ divergence at $A = 0$, a better choice of prior might be one which is 0 in $A = 0$, so as to cancel the $1/A$ factor. A convenient choice of prior which satisfies the above condition and characterizes positive parameters, is the log-normal (normalized to 1 over the range $0 \leq A \leq 1$):

$$\pi(\sigma, A) \;=\; \pi(\sigma)\,\pi(A) \;=\; H(\sigma)\,\sqrt{\frac{2}{\pi}}\,\frac{e^{-\frac{1}{2}\left(\frac{\ln A - m}{\tau}\right)^2}}{A\,\tau\,\left[1-\mathrm{erf}\left(\frac{m}{\sqrt{2}\,\tau}\right)\right]}\,H(A)\,H(1-A), \tag{7}$$

where m and τ are functions of the mean $\bar{A}$ and standard deviation ΔA of A:

$$m \;=\; \ln\frac{\bar{A}}{\sqrt{1+\left(\Delta A/\bar{A}\right)^2}} \qquad \text{and} \qquad \tau \;=\; \sqrt{\ln\left[1+\left(\Delta A/\bar{A}\right)^2\right]}. \tag{8}$$

It is easy to verify that the resulting posterior density is proper, even though the prior density is still improper with respect to σ.

Although one is usually only interested in the marginal posterior density for the cross section, it is instructive to look at the marginal posterior density for the acceptance. It can be calculated explicitly:

$$p(A\,|\,n) \;=\; \sqrt{\frac{2}{\pi}}\,e^{m-\frac{\tau^2}{2}}\,\frac{e^{-\frac{1}{2}\left(\frac{\ln A - m}{\tau}\right)^2}}{A^2\,\tau\,\left[1-\mathrm{erf}\left(\frac{m-\tau^2}{\sqrt{2}\,\tau}\right)\right]}\,H(A)\,H(1-A). \tag{9}$$

This expression has two striking aspects: it does not depend on the data n, and is a different function of A than the prior (7). No matter how weak the information contained in the data, the posterior will *never* match the prior. This lack of consistency is due to the impropriety of the prior with respect to the

cross section σ. If we regularize the prior by introducing an upper cutoff $\sigma_{\max}$ on the cross section, the posterior becomes:

$$p(A \mid n) \;\propto\; \pi(A)\, e^{-b} \left\{ \sum_{i=0}^{n} \frac{b^i - e^{-\sigma_{\max} LA}\,(\sigma_{\max} LA + b)^i}{(\sigma_{\max} LA)\, i!} \right\}, \tag{10}$$

and depends now explicitly on the data n. For low integrated luminosity L one recovers $p(A \mid n) \approx \pi(A)$, as expected.

2.2 Alternative prior 2: Gaussian in A, Gaussian in σ

Here we set the prior for (σ, A) to the product of a truncated Gaussian for σ and a truncated Gaussian for A. Such a prior is proper with respect to both A and σ. It is not a very common choice, since in general one will select a non-informative prior for the cross section. However, it provides a convenient model to study how information flows from the priors to the posterior. In particular, we want to contrast situations where the prior information agrees with the data and situations where the two are in conflict. For the examples below we set $n = 5$, $b = 2$, $A_0 = 0.02$, $\Delta A = 0.006$, and $L = 100$ pb^{-1}.

For there to be no conflict between priors and data, the cross section prior should be centered around $\sigma_0 = (n - b)/(LA_0) = 1.5$ pb. This is illustrated in the top two entries of Table 1 for two different prior uncertainties on the cross section, 30% and 5%. In both cases the measurement affects only the uncertainties, not the mean values. For the bottom two entries of the table, the cross section prior was given a mean of 7 pb in order to demonstrate the effect of a conflict between priors and data. *Both* posteriors are now shifted with respect to the priors. Comparing measurements 3 and 4, one observes that the acceptance shift increases as the prior uncertainty on the cross section is reduced.

Table 1: Summary of four measurements of a cross section, based on 5 events observed over an expected background of 2 events in a data sample of 100 pb^{-1}. The acceptance and cross section priors are truncated Gaussians. Measurements 1 and 2 illustrate the effect on the posterior of a reinforcement between data and priors, whereas measurements 3 and 4 show the effect of a conflict between the two.

		Mean		RMS/mean (%)	
Measurement		Prior	Posterior	Prior	Posterior
1:	Acceptance	0.02	0.02	30.0	27.9
	Cross section (pb)	1.5	1.5	30.0	27.9
2:	Acceptance	0.02	0.02	30.0	27.6
	Cross section (pb)	1.5	1.5	5.0	5.0
3:	Acceptance	0.020	0.014	30.0	41.3
	Cross section (pb)	7.0	4.9	30.0	41.3
4:	Acceptance	0.020	0.010	30.0	39.2
	Cross section (pb)	7.0	6.9	5.0	5.1

2.3 Summary of prior study

Our first two choices of prior, eqs. (5) and (7), show that improper priors can lead to posterior pathologies such as divergence or inconsistency. Nevertheless, improper priors are widely used as *approximations* to proper distributions representing weak information. This is justified on the grounds that "if the prior information is weak relative to the information in the data, then posterior inferences will be robust to mis-specification of the prior" (section 4.35 in [3]). From our example it would therefore appear that an improper cross section prior is not a good representation of weak information when the acceptance prior gives non-zero probability to zero acceptance.

On the other hand, it is possible to approach this problem differently, namely as the combination of two independent measurements: one of the acceptance A, summarized by the prior $\pi(A)$, and one of the expected number of signal events μ, summarized by a posterior $p(\mu \,|\, n)$. The latter is obtained via Bayes' theorem from the likelihood $(\mu+b)^n e^{-\mu-b}/n!$ and a prior $\pi(\mu)$. A simple convolution of $p(\mu \,|\, n)$ with $\pi(A)$ then yields the cross section posterior $p(\sigma \,|\, n)$. Now, it turns out that $p(\sigma \,|\, n)$ remains proper even if $\pi(A)$ is a truncated Gaussian and $\pi(\mu)$ is uniform and improper. This suggests that a different parametrization of the problem of section 2 at the prior level, in terms of μ and A instead of σ and A, would yield a more robust solution.

Our last choice of prior, in section 2.2, demonstrates how measurements can update the nuisance parameters in a way that depends not only on the data, but also on prior beliefs about the parameters of interest. This is of course the primary reason why one never cares to look at updated nuisance parameter information, even when the nuisance parameter is as *interesting* as the reusable calibration or efficiency of an instrument. If updating nuisance parameter information serves no purpose, it seems then legitimate to wonder whether it is possible to avoid this feature in a Bayesian analysis, so that all the information in the data is applied only to the parameters of interest. We look at this issue in more detail in the next section.

3 THE METHOD OF POSTERIOR AVERAGING

A salient feature of the example discussed in section 2 is that the data only depends on the *product* of the cross section and the acceptance. In this section we generalize this feature by considering measurements in which the data cannot distinguish between the parameter of interest θ and the nuisance parameter ν. In mathematical terms, we assume that there exists a function $\eta(\theta, \nu)$, independent of the data x, such that the likelihood depends on θ and ν only through η:

$$\mathcal{L}(x \,|\, \theta, \nu) \;=\; \widetilde{\mathcal{L}}(x \,|\, \eta(\theta, \nu)). \tag{11}$$

If there is more than one parameter of interest or more than one nuisance parameter, we assume that η is a vector with the same dimension as θ, and that the Jacobian of the transformation $\theta \to \eta$ is non-singular.

In order to address the last issue raised in section 2, we would like to set up the measurement problem in such a way that nuisance parameters do *not* get updated by the measurement. Thus, given a nuisance prior $\pi(\nu)$, we search for a combined prior $\pi(\theta, \nu)$ such that:

$$\int d\theta \; \pi(\theta, \nu) \;=\; \pi(\nu) \qquad \text{and} \qquad p(\nu \,|\, x) \;=\; \pi(\nu). \tag{12}$$

With the help of Bayes' theorem, the equation on the right can be rewritten as:

$$\frac{\int d\theta \; \mathcal{L}(x \,|\, \theta, \nu) \, \pi(\theta, \nu)}{\int d\nu \int d\theta \; \mathcal{L}(x \,|\, \theta, \nu) \, \pi(\theta, \nu)} \;=\; \pi(\nu). \tag{13}$$

Next, we note that under a change of variable $\theta \to \eta \equiv \eta(\theta, \nu)$, probabilities remain invariant; in particular: $\pi(\theta, \nu) \; d\theta \;=\; \pi(\eta, \nu) \; d\eta$. Equation (13) can therefore be rewritten as:

$$\frac{\int d\eta \; \widetilde{\mathcal{L}}(x \,|\, \eta) \; \pi(\eta, \nu)}{\int d\nu \int d\eta \; \widetilde{\mathcal{L}}(x \,|\, \eta) \; \pi(\eta, \nu)} \;=\; \pi(\nu), \tag{14}$$

and is satisfied by any $\pi(\eta, \nu)$ that factorizes into $\pi(\eta) \, \pi(\nu)$. Transforming back to (θ, ν), the solution is:

$$\pi(\theta, \nu) \;=\; \pi(\eta(\theta, \nu)) \; \pi(\nu) \; \frac{\partial \eta}{\partial \theta}, \tag{15}$$

where $\partial \eta / \partial \theta$ is a shorthand for the Jacobian of the transformation $\theta \to \eta$.

Using Bayes' theorem, the marginal posterior probability for θ can be written as follows:

$$p(\theta \,|\, x) \;=\; \int d\nu \; p(\theta \,|\, \nu, x) \; p(\nu \,|\, x) \;=\; \int d\nu \;\; \left[\frac{\widetilde{\mathcal{L}}(x \,|\, \eta(\theta,\nu)) \; \pi(\eta(\theta,\nu))}{\int d\eta' \; \widetilde{\mathcal{L}}(x \,|\, \eta') \; \pi(\eta')} \; \frac{\partial \eta}{\partial \theta} \right] \; \pi(\nu). \tag{16}$$

The expression between square brackets is the posterior density of η given x, written as a function of θ and ν with the help of the Jacobian $\partial\eta/\partial\theta$. The marginal posterior for θ is then the average of this η posterior over the nuisance parameter, and we will therefore refer to this method for getting rid of nuisance parameters as "posterior averaging". In a typical application of this formalism, one would identify the function $\eta(\theta,\nu)$, choose a prior for η, calculate the posterior density for η and express it in terms of θ and ν, and finally average this posterior over the prior nuisance distribution. An alternate method, appropriate when one has significant prior information about the parameter of interest θ, consists in solving the following integral equation for the η prior:

$$\pi(\theta) \;=\; \int d\nu \; \pi(\eta(\theta,\nu)) \; \pi(\nu) \; \frac{\partial \eta}{\partial \theta}. \tag{17}$$

This η prior is then handled in the same way as in the first method.

When compared with the standard method for getting rid of nuisance parameters, posterior averaging has two main advantages. First, it involves one less integration in the calculation of the marginal data density $p(x)$, making it less likely to yield divergent results when improper priors are used. Second, it exploits the fact that the data cannot distinguish between parameters of interest and nuisance parameters by channeling all the data information into the parameters of interest.

We emphasize that posterior averaging is derived from Bayes' theorem using the rules of probability without any additional, external principle, other than our method for choosing a prior. It can be applied whenever the condition embodied in eq. (11) is satisfied.

3.1 Example

We apply the posterior-averaging formalism to the example of section 2. From the likelihood (4) it is clear that the data n cannot distinguish between the cross section σ and the acceptance A. A natural choice for the function η is therefore $\eta(\sigma, A) \;=\; \sigma L A + b$, i.e. the expected number of events. An arguably non-informative prior for η is a truncated flat distribution from $\eta = b$ up to $\eta = \eta_{\max}$. For the acceptance we keep the truncated Gaussian prior of eq. (5). Using eq. (15), the combined prior for σ and A is:

$$\pi(\sigma, A) \;=\; \frac{H(\sigma) \; H(\eta_{\max} - \sigma L A - b)}{\eta_{\max} - b} \; \frac{e^{-\frac{1}{2}\left(\frac{A - A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\, K \Delta A} \; H(A) \; H(1-A) \; L A, \tag{18}$$

which does not factorize as a function of σ and A. The marginal prior for σ however, can be calculated explicitly:

$$\pi(\sigma) \;\;=\;\; C \; H(\sigma) \; \left[\frac{e^{-\left(\frac{A_0}{\sqrt{2}\Delta A}\right)^2} - e^{-\left(\frac{u(\sigma) - A_0}{\sqrt{2}\Delta A}\right)^2}}{\sqrt{\pi/2}\,(A_0/\Delta A)} + \mathrm{erf}\left(\frac{A_0}{\sqrt{2}\,\Delta A}\right) + \mathrm{erf}\left(\frac{u(\sigma) - A_0}{\sqrt{2}\,\Delta A}\right) \right], \tag{19}$$

$$\text{with} \;\; C \;\;\equiv\;\; \frac{L A_0}{(\eta_{\max} - b)\, K} \quad \text{and} \quad u(\sigma) \;\equiv\; \min\left(1, \; \frac{\eta_{\max} - b}{\sigma\, L}\right).$$

In the limit of small ΔA this expression reduces to a step function that is non-zero from $\sigma = 0$ up to $\sigma \approx (\eta_{\max} - b)/(L A_0)$. The marginal posterior density for σ is given by:

$$p(\sigma \,|\, n) \;\;=\;\; \frac{1}{S} \; H(\sigma) \; \int_0^{u(\sigma)} dA \; \frac{(\sigma A L + b)^n}{n!} \; e^{-(\sigma A L + b)} \; A L \; \frac{e^{-\frac{1}{2}\left(\frac{A - A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\, K \, \Delta A}, \tag{20}$$

$$\text{with } S \equiv \sum_{i=0}^{n} \frac{1}{i!} \left[e^{-b} b^i - e^{-\eta_{\max}} \eta_{\max}^i \right].$$

Although the priors (18) and (19) depend on the expected background b, this dependence is hardly visible for large values of $\eta_{\max}$. In any case, if this is a problem one can always choose a σ prior that is independent of b and then use eq. (17) to extract the η prior. An easier solution is to choose a flat improper prior for η by removing $(\eta_{\max} - b)$ from the denominator in eq. (18) and taking the limit $\eta_{\max} \to \infty$. The marginal prior for σ then becomes also improper, while its posterior simplifies to:

$$p(\sigma \,|\, n) = \frac{1}{S_\infty} H(\sigma) \int_0^1 dA \, \frac{(\sigma AL + b)^n}{n!} \, e^{-(\sigma AL + b)} \, AL \, \frac{e^{-\frac{1}{2}\left(\frac{A - A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\, K \, \Delta A}, \tag{21}$$

$$\text{with } S_\infty \equiv e^{-b} \sum_{i=0}^{n} \frac{b^i}{i!}.$$

Thanks to the factor A in the integrand this posterior is proper, in contrast with the calculation of section 2. In ref. [4], eq. (21) is incorrectly derived as a simple application of the technique of marginalization. The author starts with a properly normalized probability density for the parameters of interest, conditional on the nuisance parameters. This density is then averaged over the nuisance prior. However, as the first equality in eq. (16) indicates, Bayes' theorem requires that the posterior density for the parameters of interest be averaged over the nuisance posterior, not its prior. Only in special circumstances, such as those outlined at the beginning of section 3, can the two be made equal.

4 THE CONVOLUTION METHOD

In this section we briefly investigate ways to simplify computation of the posterior averaging integral of eq. (16). Often, the only prior information available about the nuisance parameter is its mean $\nu_\circ$ and standard deviation $\Delta\nu$, and a Gaussian prior is then assumed. The posterior averaging integral therefore takes the form:

$$p(\theta \,|\, x) = \int d\nu \; p(\theta \,|\, \nu, x) \, \frac{e^{-\frac{1}{2}\left(\frac{\nu - \nu_\circ}{\Delta\nu}\right)^2}}{\sqrt{2\pi}\, \Delta\nu}. \tag{22}$$

When the dependence of $p(\theta \,|\, \nu, x)$ on ν is not as easily obtainable as its dependence on θ, it is tempting to replace eq. (22) by a Gaussian convolution over θ, with some width σ to be determined as a function of $\Delta\nu$. A convolution will automatically "widen" the probability density for θ, which is the desired effect of a systematic uncertainty. To derive such a procedure, one introduces a shift function $\rho_\theta(\epsilon)$, defined as follows:

$$P(\theta \,|\, \nu_\circ + \epsilon, x) = P(\theta + \rho_\theta(\epsilon) \,|\, \nu_\circ, x), \tag{23}$$

where $P(\theta \,|\, \nu, x) \equiv \int_{-\infty}^{\theta} d\theta' p(\theta' \,|\, \nu, x)$. After performing the substitution $\nu \to \zeta \equiv \theta + \rho_\theta(\nu - \nu_\circ)$, the averaging integral (22) becomes:

$$p(\theta \,|\, x) = \int d\zeta \; p(\zeta \,|\, \nu_\circ, x) \, \frac{e^{-\frac{1}{2}\left(\frac{\zeta - \theta}{\sigma(\theta,\zeta)}\right)^2}}{\sqrt{2\pi}\, \sigma(\theta, \zeta)} \left[1 + \frac{\zeta - \theta}{\sigma(\theta,\zeta)} \frac{\partial \sigma(\theta,\zeta)}{\partial \theta} \right], \tag{24}$$

where:

$$\sigma(\theta, \zeta) \equiv \frac{\zeta - \theta}{\rho_\theta^{-1}(\zeta - \theta)} \, \Delta\nu. \tag{25}$$

Equation (24) is a good starting point for approximations. First, note that due to the Gaussian factor in eq. (22), eq. (23) only needs to hold for ϵ values of the order of a few $\Delta\nu$'s. If $\Delta\nu$ is small enough, $\rho_\theta(\epsilon)$ can be assumed to be approximately linear in ϵ, so that eq. (25) simplifies to $\sigma(\theta, \zeta) = \rho_\theta(\Delta\nu)$.

A further simplification occurs when considering a systematic uncertainty whose *main* effect is to shift the whole density $p(\theta \mid \nu, x)$ along θ without (too much) distorting its shape. In that case $\rho_\theta(\epsilon) \approx c{\cdot}\epsilon$, with c a constant independent of θ, and the posterior density becomes:

$$p(\theta \mid x) \;\approx\; \int d\zeta \; p(\zeta \mid \nu_\circ, x) \; \frac{e^{-\frac{1}{2}\left[\frac{\theta-\zeta}{c\,\Delta\nu}\right]^2}}{\sqrt{2\pi}\, c\, \Delta\nu}. \tag{26}$$

If on the other hand the systematic uncertainty under consideration mainly affects the width of $p(\theta \mid \nu, x)$, then $\rho_\theta(\epsilon) \approx -\frac{\epsilon}{\nu_\circ+\epsilon}(\theta - \theta_\circ)$, with $\theta_\circ$ the location parameter of $p(\theta \mid \nu, x)$, and:

$$p(\theta \mid x) \;\approx\; \int d\zeta \; p(\zeta \mid \nu_\circ, x) \; \frac{e^{-\frac{1}{2}\left[\frac{\theta-\zeta}{(\zeta-\theta_\circ)\Delta\nu/\nu_\circ}\right]^2}}{\sqrt{2\pi}\,(\zeta - \theta_\circ)\Delta\nu/\nu_\circ}. \tag{27}$$

Systematic uncertainties that affect both the location and width of $p(\theta \mid \nu, x)$, such as acceptance uncertainties, will give rise to non-trivial Jacobian factors in the convolution integral.

Finally, we note that in the above discussion we have ignored the transformation of integration limits. This obviously needs to be taken into account in specific applications.

Acknowledgements

I wish to thank Louis Lyons, Harrison Prosper, and Michael Goldstein for helpful discussions, James Linnemann and Louis Lyons for their comments on a first draft of this paper, and the organizers of the conference on Advanced Statistical Techniques for their hospitality.

References

[1] D. E. Groom *et al.*, "Review of Particle Physics", Eur. Phys. J. C **15**, 1 (2000).

[2] J. Linnemann, "Upper Limits and Priors", talk at the *Fermilab Workshop on Confidence Limits*, March 2000, http://conferences.fnal.gov/cl2k/.

[3] Anthony O'Hagan, "Kendall's advanced theory of statistics", Volume 2B "Bayesian inference", first edition, 1994, 330pp, Halsted Press, New York.

[4] M. Corradi, "Inclusion of systematic uncertainties in upper limits and hypothesis tests", in *Proceedings of the CERN Workshop on Confidence Limits, 17-18 January 2000*, CERN Report 2000-005.