BAYESIAN TREATMENTS OF SYSTEMATIC UNCERTAINTIES

Luc Demortier

The Rockefeller University

Contents:

- Nuisance Parameter Treatment
- A difficult case: Cross Section Measurement with acceptance uncertainties
- Solutions based on Prior Selection:
  - Log-normal prior in A, flat prior in σ_s
  - Gaussian prior in A, Gaussian prior in σ_s
  - Gaussian prior in A, flat prior in σ_s
- Solution based on Posterior Averaging
- The Convolution Method

NUISANCE PARAMETER TREATMENT

Given:
- data x
- parameter of interest θ
- nuisance parameter ν
- Likelihood $\mathcal{L}(x|\theta,\nu)$
- Prior $\pi(\theta,\nu)$

The Posterior Probability Density is:

$$p(\theta,\nu|x) = \frac{\mathcal{L}(x|\theta,\nu)\ \pi(\theta,\nu)}{\int d\theta \int d\nu\ \mathcal{L}(x|\theta,\nu)\ \pi(\theta,\nu)}$$

To obtain a density for θ only, integrate out ν:

$$p(\theta|x) = \frac{\int d\nu\ \mathcal{L}(x|\theta,\nu)\ \pi(\theta,\nu)}{\int d\theta \int d\nu\ \mathcal{L}(x|\theta,\nu)\ \pi(\theta,\nu)}$$

"Marginal posterior density for θ"

Very often the prior will be factorizable:

$$\pi(\theta, \nu) = \pi_1(\theta) \cdot \pi_2(\nu)$$

in which case the marginal density for θ can be rewritten as:

$$p(\theta \,|\, x) = \frac{\mathcal{L}'(x|\theta)\ \ \pi_1(\theta)}{\int d\theta\ \mathcal{L}'(x|\theta)\ \pi_1(\theta)}$$

where:

$$\mathcal{L}'(x|\theta) \equiv \int d\nu\ \mathcal{L}(x\,|\,\theta, \nu)\ \ \pi_2(\nu)$$

i.e. apply Bayes' theorem directly to a "modified" likelihood.

Cross Section Measurement with Acceptance Uncertainties

We apply the nuisance parameter treatment to the measurement of:

a signal cross section σ_s

in the presence of:

a number of observed events n

an expected background b

an acceptance A

an integrated luminosity L

The likelihood is Poisson:

$$\mathcal{L}(n \mid \sigma_s, A) = \frac{1}{n!} (\sigma_s L A + b)^n e^{-(\sigma_s L A + b)}$$

For the prior we take:

$$\pi(\sigma_s, A) = \pi_1(\sigma_s)\, \pi_2(A) = \frac{e^{-\frac{1}{2}\left(\frac{A - A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\, K\, \Delta A}$$

$$(\sigma_s \geqslant 0 \text{ and } 0 \leqslant A \leqslant 1)$$

Note: this prior is improper w.r.t. σ_s !

The marginal posterior density for σ_s is then:

$$p(\sigma_s \mid n) = \frac{1}{C} \int_0^1 dA \, \frac{(\sigma_s L A + b)^n}{n!} e^{-(\sigma_s L A + b)} \frac{e^{-\frac{1}{2}\left(\frac{A-A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\, K \Delta A}$$

where:

$$C = \int_0^{+\infty} d\sigma_s \int_0^1 dA \, \frac{(\sigma_s L A + b)^n}{n!} e^{-(\sigma_s L A + b)} \frac{e^{-\frac{1}{2}\left(\frac{A-A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\, K \Delta A}$$

$$= \frac{e^{-b}}{L} \sum_{i=0}^{n} \frac{b^i}{i!} \int_0^1 dA \, \frac{1}{A} \frac{e^{-\frac{1}{2}\left(\frac{A-A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\, K \Delta A}.$$

$$= \infty$$

⇒ The posterior is improper

⇒ It's impossible to extract upper limits ...

NOTE!

The problem is not due to the choice of a Gaussian for the acceptance prior but to:

1. A flat cross section prior
2. An acceptance prior that gives non-zero probability to zero acceptance.

Before suggesting a "quick" solution to this problem, remember Cromwell's dictum, which admonishes not to give zero probability to any event that is possible, however highly improbable.

. . . .

Solutions based on prior selection

① Log-normal prior in A, flat prior in σ_s.

Try:

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

$$(\sigma_s \geq 0 \quad \text{and} \quad 0 \leq A \leq 1)$$

where:

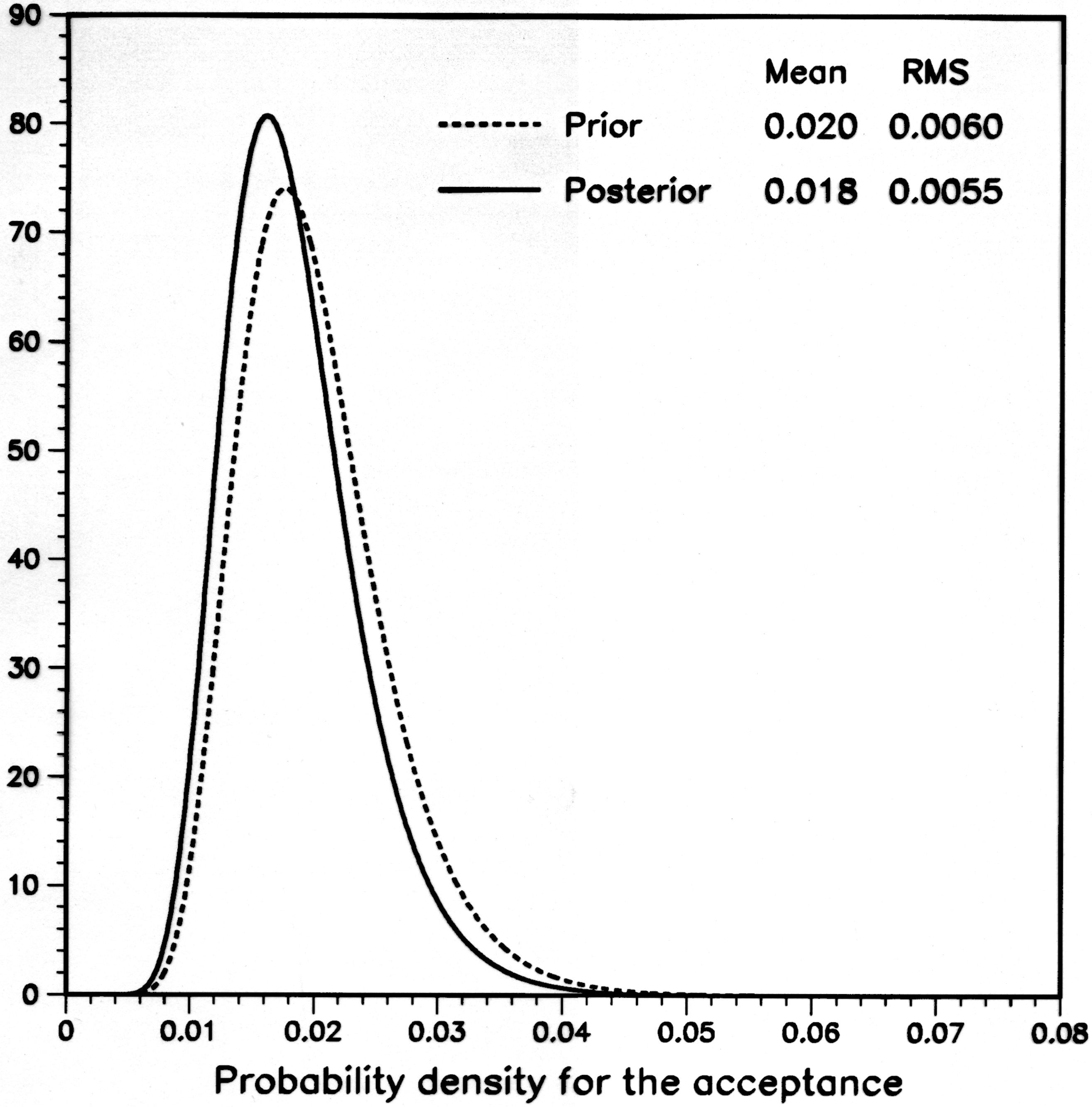
$$m = \ln \frac{\bar{A}}{\sqrt{1 + \left(\frac{\Delta A}{\bar{A}}\right)^2}} \quad \text{and} \quad \tau = \sqrt{\ln\left[1 + \left(\frac{\Delta A}{\bar{A}}\right)^2\right]}$$

Note that $\pi(\sigma_s, A) \to 0$ as $A \to 0$.

$\Rightarrow$ Posterior is proper !!

F.e. look at marginal posterior for A:

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



Mean RMS
Prior 0.020 0.0060
Posterior 0.018 0.0055
90
80
70
60
50
40
30
20
10
0
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
Probability density for the acceptance

Solutions based on prior selection

② Gaussian prior in A, Gaussian prior in σ_s:

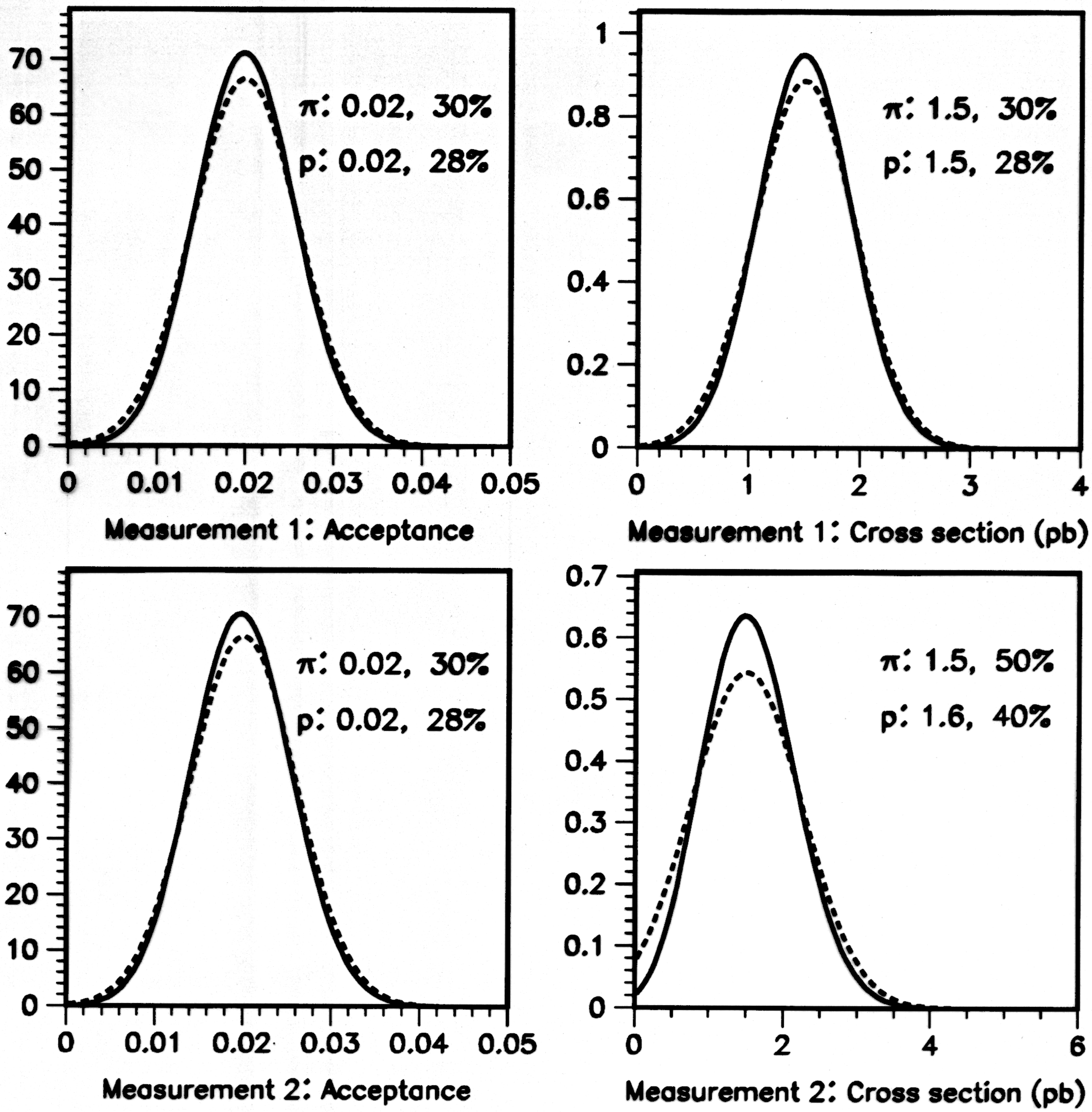
$$\pi(\sigma_s, A) = \frac{e^{-\frac{1}{2}\left(\frac{\sigma_s - \sigma_{s0}}{\Delta\sigma_s}\right)^2}}{\sqrt{2\pi}\, K_1\, \Delta\sigma_s} \; \frac{e^{-\frac{1}{2}\left(\frac{A - A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\, K_2\, \Delta A}$$

$$(\sigma_s \geq 0 \quad \text{and} \quad 0 \leq A \leq 1)$$

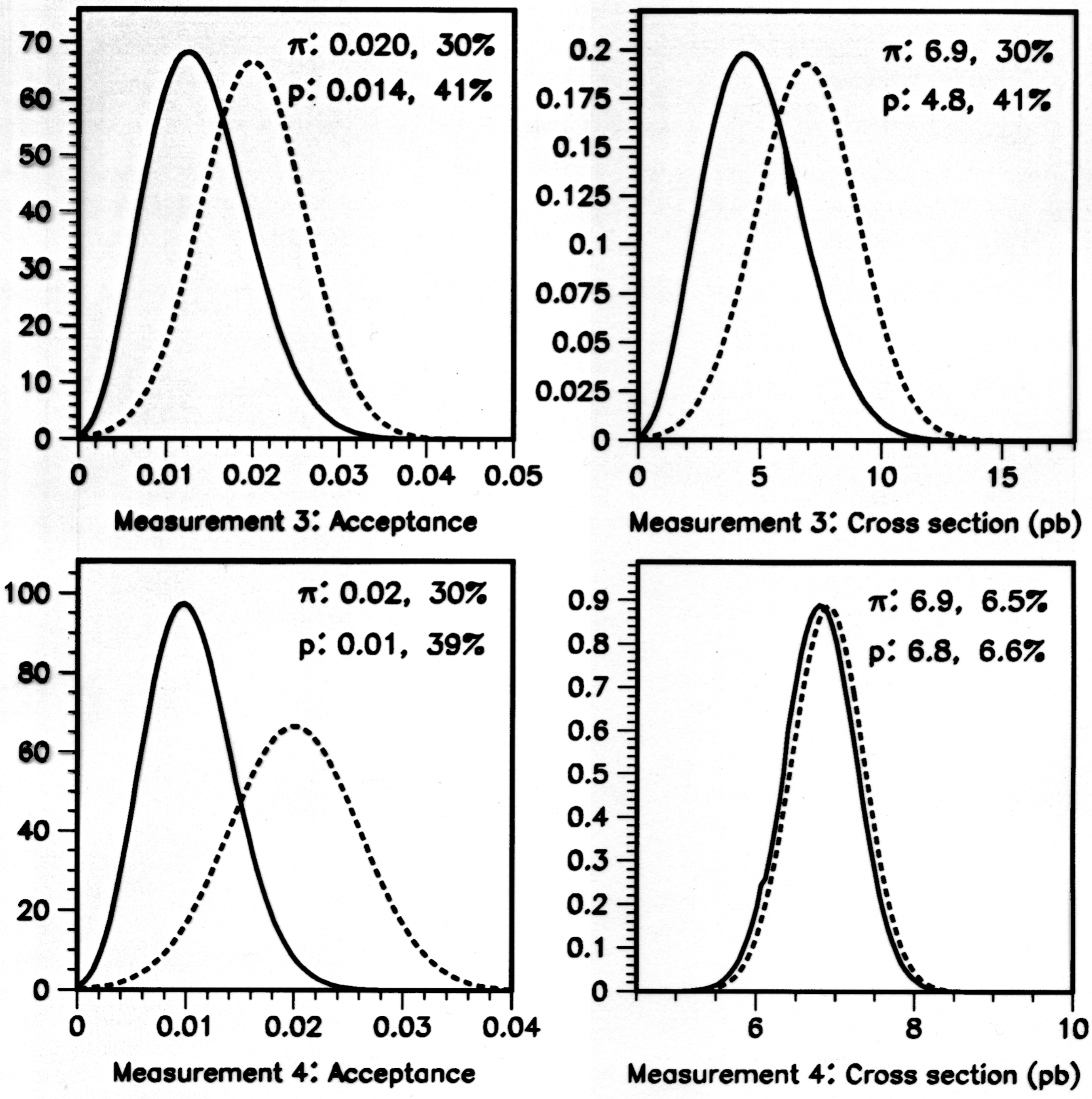
Prior is proper $\Rightarrow$ Posterior is proper.

Examples (for $n = 5$, $b = 2$, $L = 100\ pb^{-1}$):

		Mean Prior	Mean Post.	RMS/Mean (%) Prior	RMS/Mean (%) Post.
1	A	0.02	0.02	30	28
	σ_s	1.5 pb	1.5 pb	30	28
2	A	0.02	0.02	30	28
	σ_s	1.5 pb	1.55 pb	50	40
3	A	0.020	0.014	30	41
	σ_s	6.9 pb	4.8 pb	30	41
4	A	0.020	0.010	30	39
	σ_s	6.9 pb	6.8 pb	6.5	6.6



π: 0.02, 30%
p: 0.02, 28%
Measurement 1: Acceptance
π: 1.5, 30%
p: 1.5, 28%
Measurement 1: Cross section (pb)
π: 0.02, 30%
p: 0.02, 28%
Measurement 2: Acceptance
π: 1.5, 50%
p: 1.6, 40%
Measurement 2: Cross section (pb)


π: 0.020, 30%
p: 0.014, 41%
Measurement 3: Acceptance
π: 6.9, 30%
p: 4.8, 41%
Measurement 3: Cross section (pb)
π: 0.02, 30%
p: 0.01, 39%
Measurement 4: Acceptance
π: 6.9, 6.5%
p: 6.8, 6.6%
Measurement 4: Cross section (pb)

Solutions based on prior selection

③ Gaussian prior in A, truncated flat prior in σ_s.

$$\pi(\sigma_s, A) = \frac{1}{\sigma_{max}} \frac{e^{-\frac{1}{2}\left(\frac{A-A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\, K\, \Delta A}$$

$$(0 \leq \sigma_s \leq \sigma_{max} \text{ and } 0 \leq A \leq 1)$$

The posterior normalization is:

$$C = \int_0^1 dA \, \frac{e^{-\frac{1}{2}\left(\frac{A-A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\, K\, \Delta A} \, \frac{e^{-b}}{L\, \sigma_{max}} \sum_{i=0}^{n} \frac{b^i - (\sigma_{max} L A + b)\, e^{-\sigma_{max} L A}}{A\, i!}$$

Note that:

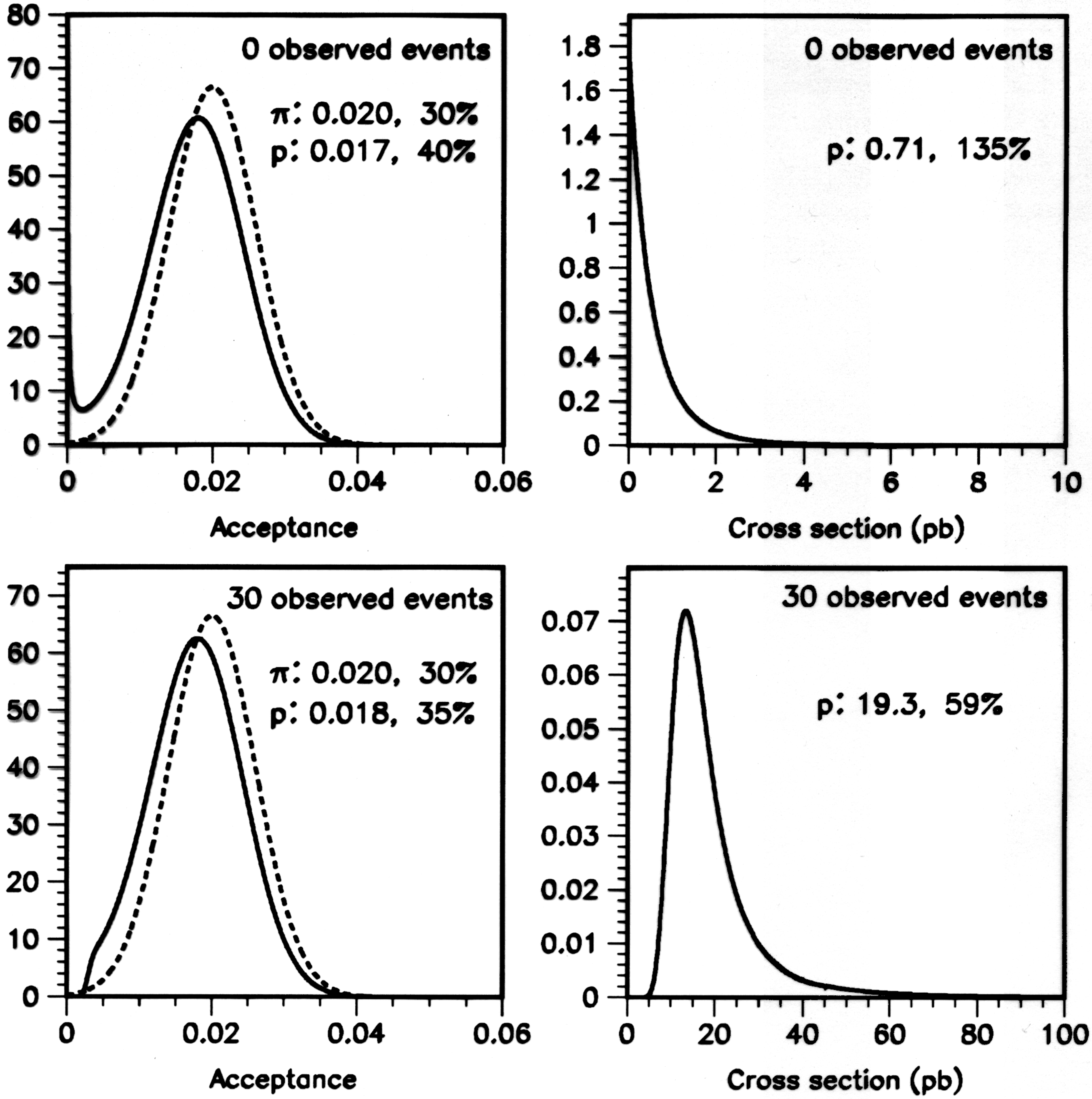
$$\lim_{A \to 0} \sum_{i=0}^{n} \frac{b^i - (\sigma_{max} L A + b)\, e^{-\sigma_{max} L A}}{A\, i!} = \frac{b^n}{n!} L\, \sigma_{max}$$

The posterior density is therefore proper.

Examples for $\sigma_{max} = 100\,pb$

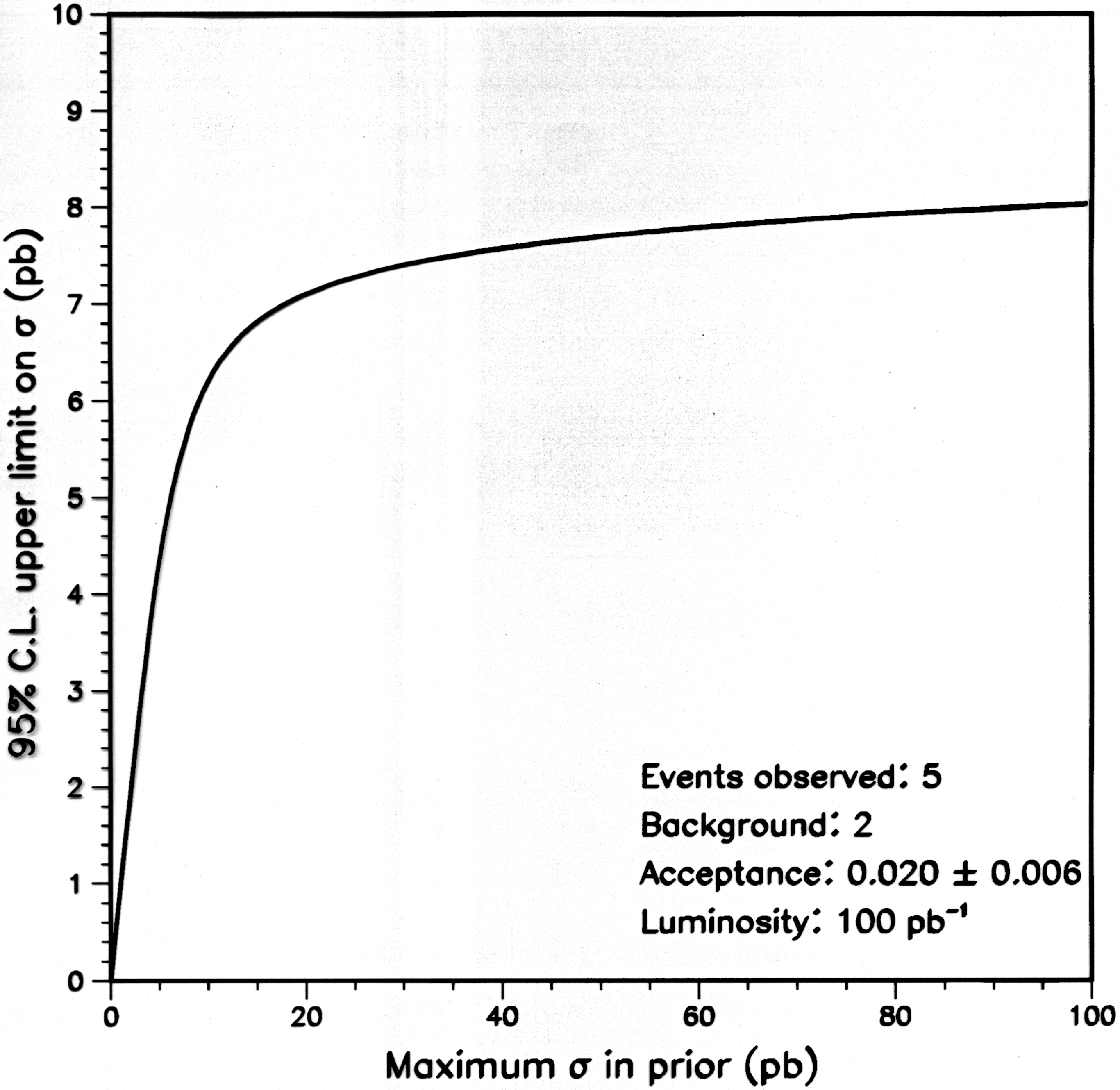
$A_0 = 0.02 \qquad \Delta A = 0.006$

$b = 2 \qquad L = 100\,pb^{-1}$



$n = 0$

$n = 30$



95% C.L. upper limit on σ (pb)
10
9
8
7
6
5
4
3
2
1
0
0
20
40
60
80
100
Maximum σ in prior (pb)
Events observed: 5
Background: 2
Acceptance: 0.020 ± 0.006
Luminosity: 100 pb⁻¹

Summary of Prior Study

F1 The marginalization treatment of systematic uncertainties sometimes leads to improper posteriors, as in the case of a cross section measurement with a flat (improper) cross section prior and a Gaussian acceptance prior. The form of the likelihood requires one to select a prior that is either proper w.r.t. σ_S or zero in $A=0$. Either choice may be "inconvenient".

F2 When the data cannot distinguish between the nuisance parameters and the parameters of interest, the amount of data information that goes into updating each of the corresponding priors is completely determined by the relative strengths of the priors:

Stronger prior $\Rightarrow$ smaller update

Q1 For true nuisance parameters that cannot be distinguished from the parameters of interest by the data, why don't we try to avoid updating the information about them and instead try to maximize the updating of information about the parameters of interest?

Q2 For nuisance parameters that have intrinsic interest (e.g. an energy scale or a tracking efficiency that is used repeatedly in many independent measurements), why don't we ever calculate their marginalized posterior density and use it in subsequent measurements instead of reusing the same prior over and over again? Does the answer depend on whether the nuisance parameter information was degraded due to conflict between data and prior, or reinforced due to agreement between them?

Treatment of systematic uncertainties by posterior averaging

Assume the prior factorizes:

$$\pi(\theta, \nu) = \pi(\theta) \cdot \pi(\nu)$$

Then:

Step 1: Calculate the posterior for the parameters of interest at a fixed value of the nuisance par.:

$$p(\theta \mid x, \nu) = \frac{\mathcal{L}(x \mid \theta, \nu)\ \pi(\theta)}{\int d\theta\ \mathcal{L}(x \mid \theta, \nu)\ \pi(\theta)}$$

"Posterior density for θ, conditional on ν"

Step 2: Multiply by a prior for ν and integrate out ν:

$$p(\theta \mid x) = \int d\nu\ p(\theta \mid x, \nu)\ \pi(\nu)$$

[Note: this formula obeys the rules of probability theory.]

Compare posterior average:

$$p(\theta|x) = \int d\nu \left[\frac{\mathcal{L}(x|\theta,\nu)\,\pi(\theta)}{\int d\theta\, \mathcal{L}(x|\theta,\nu)\,\pi(\theta)} \right] \pi(\nu)$$

with standard marginalization:

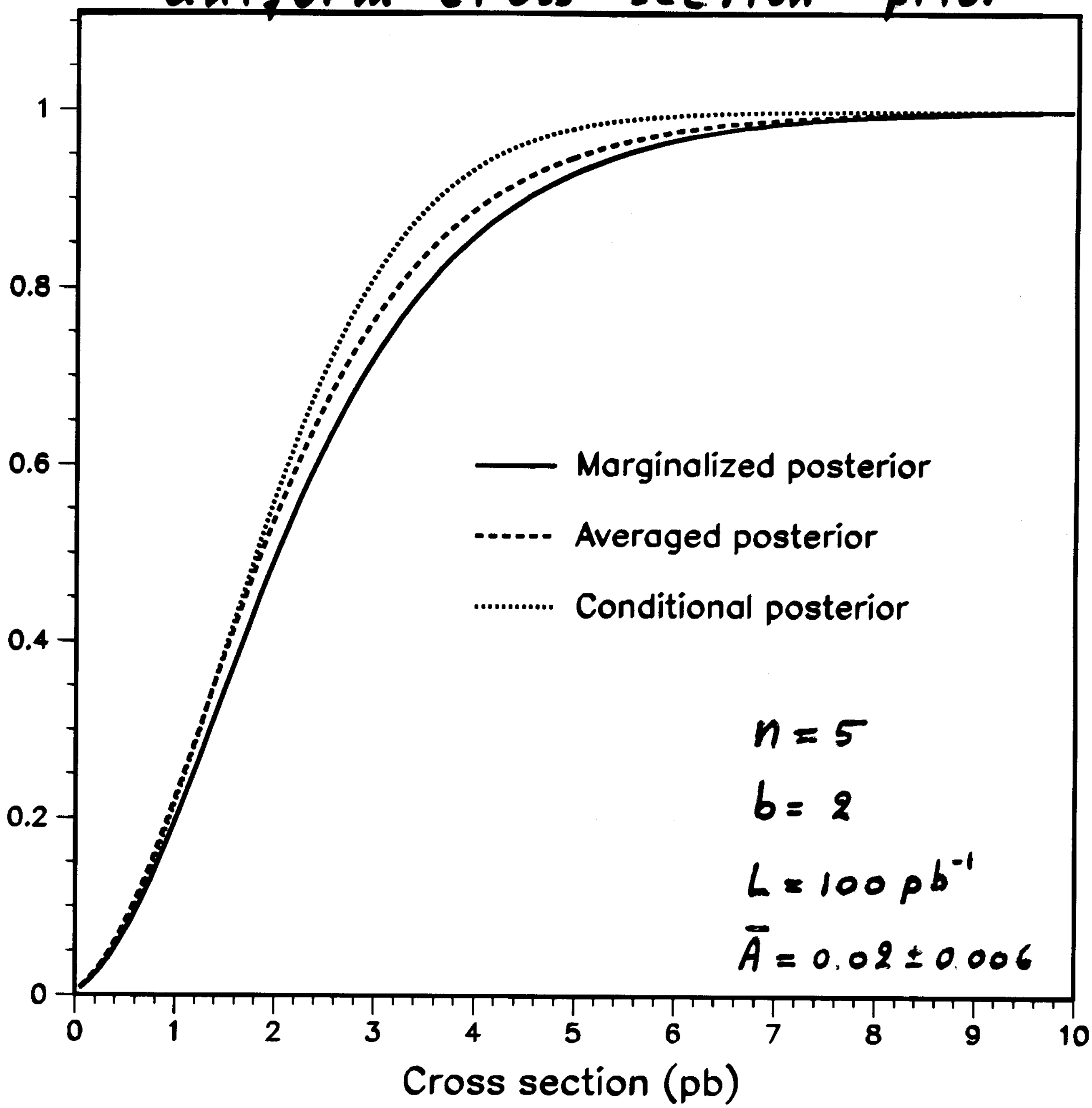
$$p(\theta|x) = \frac{\left[\int d\nu\, \mathcal{L}(x|\theta,\nu)\,\pi(\nu)\right] \pi(\theta)}{\int d\theta \int d\nu\, \mathcal{L}(x|\theta,\nu)\,\pi(\nu)\,\pi(\theta)}$$

Advantages of posterior averaging:

(1) Converges more easily, especially in cases where data cannot distinguish between nuisance parameters and parameters of interest.

(2) Does not waste data on updating nuisance parameters.

Posterior probability for the cross section
Log-normal acceptance prior
Uniform cross section prior



Posterior averaging in other contexts

(1) Talk by M. Corradi at the CERN workshop on confidence limits, January 2000.

(2) Bayesian Model Averaging:

$$p(\Delta \mid D) = \sum_{k=1}^{K} p(\Delta \mid M_k, D)\, p(M_k \mid D)$$

$$\begin{cases} \Delta = \text{quantity of interest} \\ D = \text{data} \\ M_k,\ k=1 \ldots K : \quad \text{list of models} \end{cases}$$

(3) Frequentists sometimes use posterior averaging to fold systematic uncertainties into probability densities. This frequentist-Bayesian hybrid is justified on the grounds that the systematics are small compared to the statistical uncertainties, and that the latter are treated in a proper frequentist way (Barlow).

The Convolution Method

Systematic uncertainties are sometimes incorporated into a probability density $f(\theta;\nu)$ by means of a convolution:

$$f^{\dagger}(\theta) \equiv \int d\eta \; f(\eta;\nu_0)\, \frac{e^{-\frac{1}{2}\left(\frac{\theta-\eta}{\sigma}\right)^2}}{\sqrt{2\pi}\,\sigma}$$

Advantages of this method:

(1) Guaranteed "broadening" of f;

(2) Multiple systematics add up in quadrature;

(3) Simple to calculate.

Disadvantage:

(1). How to calculate σ?

- How is σ related to $\Delta\nu$?
- How do we know a Gaussian smearing factor is appropriate?

To eliminate this disadvantage, let us try to relate the convolution method to the posterior averaging method:

$$f^*(\theta) \equiv \int d\nu \; f(\theta;\nu) \; \frac{e^{-\frac{1}{2}\left(\frac{\nu-\nu_0}{\Delta\nu}\right)^2}}{\sqrt{2\pi}\;\Delta\nu}$$

?

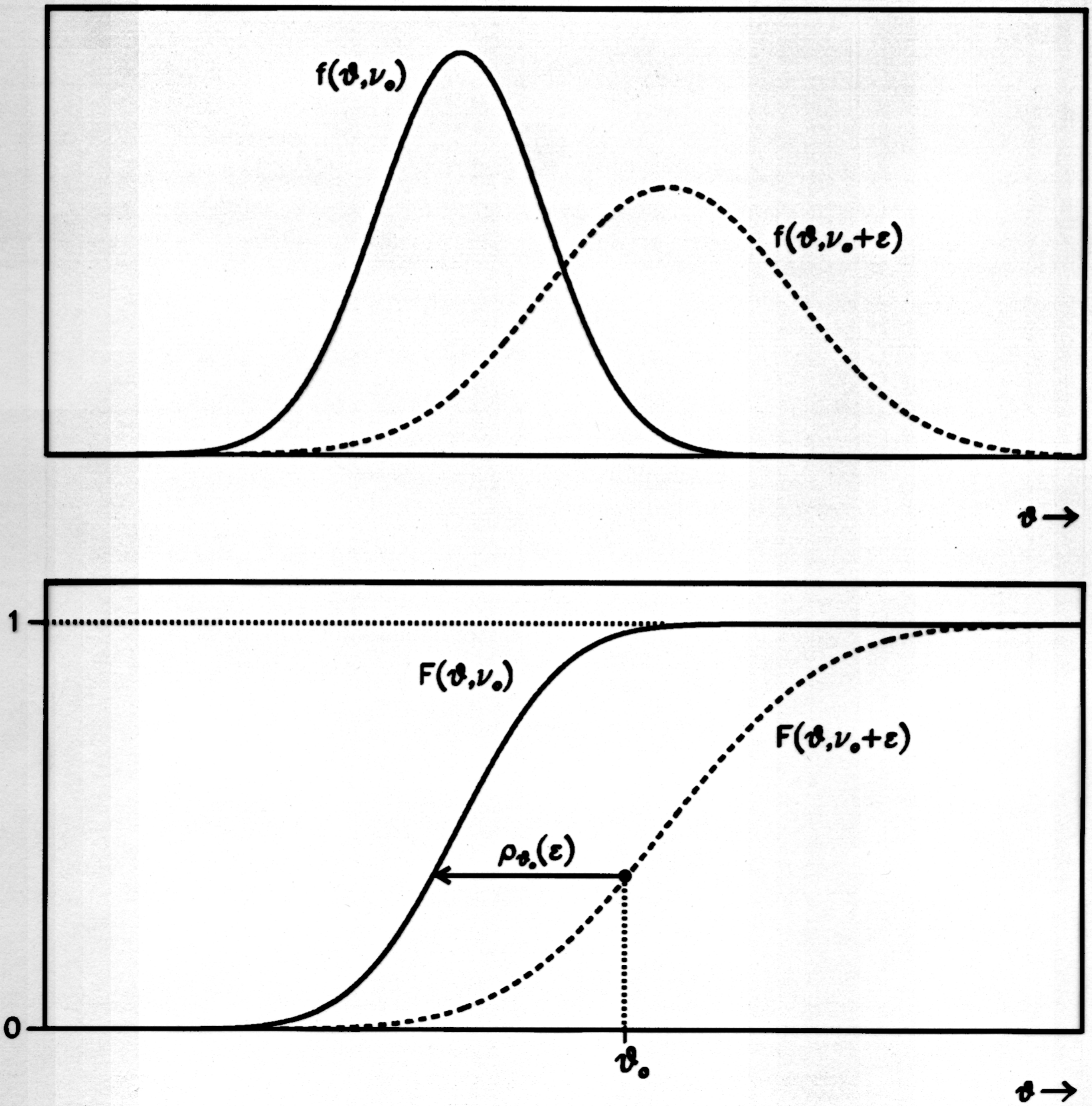
$$f^\dagger(\theta) \equiv \int d\eta \; f(\eta;\nu_0) \; \frac{e^{-\frac{1}{2}\left(\frac{\theta-\eta}{\sigma}\right)^2}}{\sqrt{2\pi}\;\sigma}$$

We need to be able to relate an arbitrary shift in ν to a shift in θ ; this can be done with the help of a shift function $\rho_\theta(\epsilon)$:

$$F(\theta;\nu_0+\epsilon) = F(\theta+\rho_\theta(\epsilon);\nu_0)$$

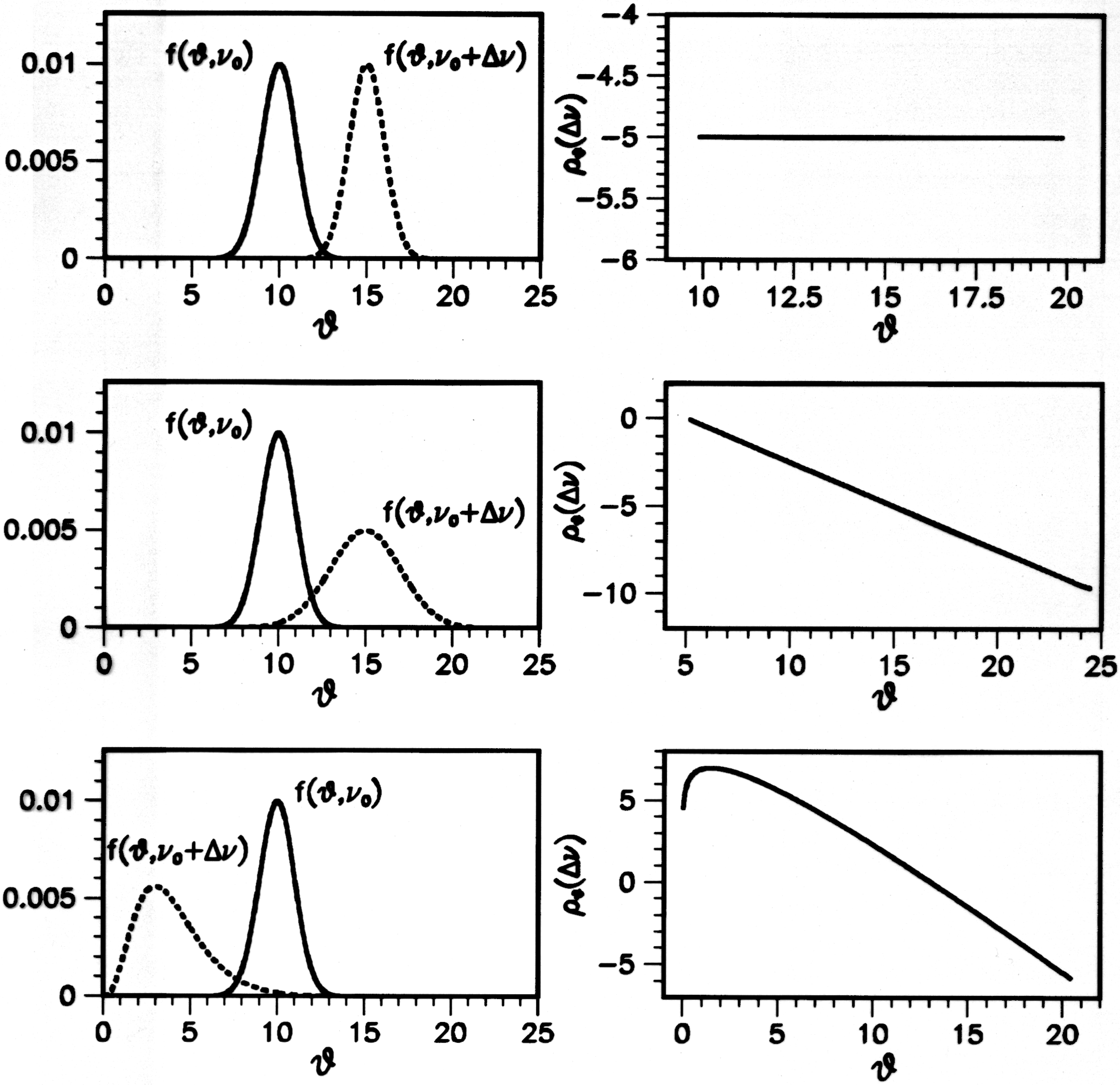
where F is the cumulative probability corresponding to f.

Construction of the shift function $\rho_\theta(\epsilon)$



$$F(\theta;\nu_0+\epsilon) = F(\theta+\rho_\theta(\epsilon);\nu_0)$$

Examples of shift functions



Starting from the expression for posterior averaging, and using the shift function, we have:

$$f^*(\theta) = \int d\epsilon \; f(\theta; \nu_0+\epsilon) \; \frac{e^{-\frac{1}{2}\left(\frac{\epsilon}{\Delta\nu}\right)^2}}{\sqrt{2\pi}\;\Delta\nu}$$

$$= \int d\epsilon \; f(\theta+\rho_\theta(\epsilon); \nu_0) \left| 1 + \frac{\partial \rho_\theta(\epsilon)}{\partial \theta} \right| \frac{e^{-\frac{1}{2}\left(\frac{\epsilon}{\Delta\nu}\right)^2}}{\sqrt{2\pi}\;\Delta\nu}$$

$$= \int d\eta \; f(\eta; \nu_0) \left| \frac{1 + \frac{\partial \rho_\theta(\epsilon)}{\partial \theta}}{\frac{\partial \rho_\theta(\epsilon)}{\partial \epsilon}} \right|_{\epsilon=\rho_\theta^{-1}(\eta-\theta)} \frac{e^{-\frac{1}{2}\left(\frac{\rho_\theta^{-1}(\eta-\theta)}{\Delta\nu}\right)^2}}{\sqrt{2\pi}\;\Delta\nu}$$

where
$$\begin{cases} \eta \equiv \theta + \rho_\theta(\epsilon) \\ \rho_\theta^{-1}(z) = \epsilon \iff \rho_\theta(\epsilon) = z \end{cases}$$

Next, define σ by:

$$\sigma(\theta, \eta) \equiv \frac{\eta-\theta}{\rho_\theta^{-1}(\eta-\theta)}\,\Delta\nu$$

Then:

$$f^*(\theta) = \int d\eta \; f(\eta; \nu_0) \, \frac{e^{-\frac{1}{2}\left(\frac{\eta-\theta}{\sigma}\right)^2}}{\sqrt{2\pi}\,\sigma} \left| 1 + \frac{\eta-\theta}{\sigma} \frac{\partial \sigma}{\partial \theta} \right|$$

This is like a convolution, except that σ depends on η and θ, and there is a Jacobian factor in the integrand.

One also needs to keep track of the integration limits.

In general however, the main advantages of the convolution method are lost. But... it can be used as a starting point for various approximations.

Linear approximation

Only values of ϵ that are of the order of a few $\Delta\nu$ contribute to the smearing integral. Assume the shift function is linear in ϵ over such a range:

$$\rho_\theta(\epsilon) = \rho'_\theta \cdot \epsilon$$

Then we find for σ :

$$\sigma(\theta,\eta) = \rho'_\theta \,\Delta\nu = \rho_\theta(\Delta\nu)$$

so that:

$$f^*(\theta) = \int d\eta \; f(\eta;\nu_0)\, \frac{e^{-\frac{1}{2}\left(\frac{\eta-\theta}{\rho'_\theta \Delta\nu}\right)^2}}{\sqrt{2\pi}\,\rho'_\theta\,\Delta\nu} \left| 1 + \frac{\eta-\theta}{\rho'_\theta} \frac{d\rho'_\theta}{d\theta} \right|$$

Further approximations will address the dependence of ρ'_θ on θ .

Translation approximation

Suppose the main effect of a systematic uncertainty is to shift the location of the probability density f:

$$\rho_\theta(\epsilon) \cong c \cdot \epsilon$$

Then:

$$f^*(\theta) \cong f^+(\theta) \cong \int d\eta \; f(\eta; \nu_0) \; \frac{e^{-\frac{1}{2}\left(\frac{\theta-\eta}{c\,\Delta\nu}\right)^2}}{\sqrt{2\pi}\; c\,\Delta\nu}$$

and all such systematics add up in quadrature.

Scaling approximation

Suppose the main effect of a systematic uncertainty is to change the width of the probability density f. Then:

$$F(\theta; \nu) \cong F(\theta \cdot \nu)$$

and

$$\rho_\theta(\epsilon) \cong \theta \frac{\epsilon}{\nu_0}$$

so that:

$$f^*(\theta) \cong \int d\eta \; f(\eta; \nu_0) \frac{e^{-\frac{1}{2}\left(\frac{\theta - \eta}{\theta \Delta\nu/\nu_0}\right)^2}}{\sqrt{2\pi}\,\theta\,\Delta\nu/\nu_0} \frac{\eta}{\theta}$$

Note the Jacobian factor η/θ.

Here too, the $\theta\,\Delta\nu/\nu_0$'s for multiple independent systematic uncertainties add up in quadrature, provided the $\Delta\nu/\nu_0$'s are small.

Combining systematic uncertainties of the translation and scaling type.

$$f^*(\theta) = \int d\eta \; f(\eta) \frac{e^{-\frac{1}{2}\left(\frac{\eta-\theta}{\sqrt{\Delta\mu^2+(\theta\delta\nu)^2}}\right)^2}}{\sqrt{2\pi}\sqrt{\Delta\mu^2+(\theta\delta\nu)^2}} \; \frac{1}{\theta} \; \frac{\frac{\eta}{\Delta\mu^2}+\frac{\theta}{(\theta\delta\nu)^2}}{\frac{1}{\Delta\mu^2}+\frac{1}{(\theta\delta\nu)^2}}$$

where $\Delta\mu$ is the σ for the combined translation uncertainty and $\theta\delta\nu$ is the σ for the combined scale uncertainty.

Summary

- We propose to handle systematic uncertainties by the method of **posterior averaging**, as this method has good convergence properties and does not update nuisance information.

- Posterior averaging integrals can be approximated by **convolutions** provided the systematic uncertainties have a small effect on the p.d.f.