Bayesian Reference Analysis

Luc Demortier

Rockefeller HEP Group Meeting

November 7, 2005

<u>Flow of ideas in this talk:</u>

- Reminder of Bayesian Paradigm;
- Reference Analysis;
- Intrinsic Discrepancy and Missing Information;
- Reference Priors;
- Example: Cross section measurements;
- Intrinsic Estimates and Credible Regions;
- Example: Transverse Momentum Measurements;
- Summary

Reminder of Bayesian Paradigm

- Suppose we have observed some data x with probability density $f(x \,|\, \theta, \nu)$.
- The likelihood $\mathcal{L}(\theta, \nu \,|\, x)$ is proportional to $f(x \,|\, \theta, \nu)$, considered as a function of (θ, ν).
- In Bayesian statistics, θ and ν are viewed as random variables, not in the sense that their true values can fluctuate, but rather that information about them can be represented coherently by a probability distribution.
- It is therefore possible to introduce a *prior* density $\pi(\theta, \nu)$, which summarizes our knowledge about (θ, ν) *prior* to the measurement x.
- Bayes' theorem specifies how the prior information about (θ, ν) is updated by the measurement of x to yield the posterior density of (θ, ν):

$$
\begin{aligned}
p(\theta, \nu) &= \mathcal{L}(\theta, \nu \,|\, x)\; \pi(\theta, \nu) \;/\; p(x) \\
\text{where:} \quad p(x) &= \int d\nu \int d\theta\; \mathcal{L}(\theta, \nu \,|\, x)\; \pi(\theta, \nu).
\end{aligned}
$$

Motivation

Those who doubt that the Bayesian paradigm has anything to offer to physicists are encouraged to read the following paper:

Carlton M. Caves, Christopher A. Fuchs, and Rüdiger Schack, "Quantum probabilities as Bayesian probabilities," Phys. Rev. **A** 65, 022305 (2002).

The probabilities of quantum theory are fully determined by physical laws, but they are subjective in the sense that their form depends on the information available to the observer.

When dealing with actual measurements however, no fundamental law constrains their analysis, summary and report, so that some other objective method must be found.

Reference Analysis

Reference analysis is a method to produce inferences that only depend on the model assumed and the data observed. It is meant to provide standards for scientific communication.

In order to be generally and consistently applicable, reference analysis uses the Bayesian paradigm, which immediately raises the question of what kind of prior will produce "objective" inferences? The primary aim is to obtain posterior distributions that are dominated in some sense by the information contained in the data, but there are additional requirements that may reasonably be considered necessary properties of any proposed solution:

- *Generality:* The procedure should be completely general and should always yield *proper* posteriors.
- *Invariance:* If $\phi = \phi(\theta)$, then $\pi(\phi \,|\, x) = \pi(\theta \,|\, x)\,|d\theta/d\phi|$. Furthermore, if $t = t(x)$ is a sufficient statistic, then $\pi(\theta \,|\, x) = \pi(\theta \,|\, t)$.
- *Consistent Marginalization* If $p(x \,|\, \theta, \lambda) \to \pi_1(\theta \,|\, x) = \pi_1(\theta \,|\, t)$, $t = t(x)$ and $p(t \,|\, \theta, \lambda) = p(t \,|\, \theta) \to \pi_2(\theta \,|\, t)$, Then $\pi_2(\theta \,|\, t) = \pi_1(\theta \,|\, t)$.
- *Consistent sampling properties.* The family of posterior distributions $\pi(\theta \,|\, x)$ obtained by repeated sampling from the model $p(x \,|\, \theta, \lambda)$ should concentrate on a region of Θ which contains the true value of θ.

Intrinsic Discrepancy (1)

Reference analysis techniques are based on information theory, and in particular on the central concept of intrinsic discrepancy between probability densities:

The intrinsic discrepancy between two probability densities p_1 and p_2 is:

$$\delta\{p_1, p_2\} \;=\; \min\left\{\int dx\, p_1(x)\;\ln\frac{p_1(x)}{p_2(x)},\;\; \int dx\, p_2(x)\;\ln\frac{p_2(x)}{p_1(x)}\right\},$$

provided one of the integrals is finite. The intrinsic discrepancy between two parametric models for x,
$\mathcal{M}_1 = \{p_1(x\,|\,\phi), x \in \mathcal{X}, \phi \in \Phi\}$ and $\mathcal{M}_2 = \{p_2(x\,|\,\psi), x \in \mathcal{X}, \psi \in \Psi\}$,
is the minimum intrinsic discrepancy between their elements:

$$\delta\{\mathcal{M}_1, \mathcal{M}_2\} \;=\; \inf_{\phi,\psi}\; \delta\{p_1(x\,|\,\phi), p_2(x\,|\,\psi)\}.$$

Intrinsic Discrepancy (2)

Properties of the intrinsic discrepancy:

- $\delta\{p_1, p_2\}$ is symmetric, non-negative, and vanishes if and only if $p_1(x) = p_2(x)$ almost everywhere.
- $\delta\{p_1, p_2\}$ is invariant under one-to-one transformations of x.
- $\delta\{p_1, p_2\}$ is information-additive: the discrepancy for a set of n independent observations is n times the discrepancy for one observation.
- The intrinsic discrepancy $\delta\{\mathcal{M}_1, \mathcal{M}_2\}$ is the minimum expected log-likelihood ratio in favor of the model which generates the data.
- The intrinsic discrepancy $\delta\{\mathcal{M}_1, \mathcal{M}_2\}$ between two parametric families of distributions does not depend on their parametrizations.
- The intrinsic discrepancy $\delta\{p_1, p_2\}$ is a measure, in natural information units, of the minimum amount of expected information required to discriminate between p_1 and p_2.

Missing Information

The expected intrinsic information $I\{p(\theta)\,|\,\mathcal{M}\}$ from one observation of

$$\mathcal{M} \equiv \{p(x\,|\,\theta),\; x \in \mathcal{X},\; \theta \in \Theta\}$$

about the value of θ when the prior density is $p(\theta)$, is:

$$I\{p(\theta)\,|\,\mathcal{M}\} \;=\; \delta\{p(x,\theta),\; p(x)\,p(\theta)\},$$

where $p(x,\theta) = p(x\,|\,\theta)\,p(\theta)$ and $p(x) = \int d\theta\; p(x\,|\,\theta)\,p(\theta)$.

The stronger the prior knowledge described by $p(\theta)$, the smaller the information the data may be expected to provide. Conversely, weak initial knowledge about θ corresponds to large expected information from the data.

Consider the intrinsic information about θ, $I\{p(\theta), \mathcal{M}^k\}$, which could be expected from making k independent observations from $\mathcal{M}$. As k increases, the true value of θ would become precisely known. Thus, as $k \to \infty$, $I\{p(\theta), \mathcal{M}^k\}$ measures the amount of *missing information* about θ which corresponds to the prior $p(\theta)$.

Reference Priors for One-Parameter Models

Let $\mathcal{P}$ be a class of sufficiently regular priors that are compatible with whatever initial information one has about the value of θ.

The reference prior is then defined to be that prior function $\pi(\theta) = \pi(\theta \,|\, \mathcal{M}, \mathcal{P})$ which maximizes the missing information about the value of θ within the class $\mathcal{P}$ of candidate priors.

Formally:

$\pi(\theta) = \pi(\theta \,|\, \mathcal{M}, \mathcal{P})$ is a reference prior for model $\mathcal{M}$ given $\mathcal{P}$ if, for some increasing sequence $\{\Theta_i\}_{i=1}^{\infty}$ with $\lim_{i\to\infty} \Theta_i = \Theta$ and $\int_{\Theta_i} \pi(\theta)\, d\theta < \infty$,

$$\lim_{k\to\infty} \left[I\{\pi_i \,|\, \mathcal{M}^k\} - I\{p_i \,|\, \mathcal{M}^k\} \right] \geq 0 \qquad \text{for all } \Theta_i, \ \text{for all } p \in \mathcal{P},$$

where $\pi_i(\theta)$ and $p_i(\theta)$ are the renormalized restrictions of $\pi(\theta)$ and $p(\theta)$ to Θ_i.

Some Properties of Reference Priors

- In the definition, the limit $k \to \infty$ is *not* an approximation, but an essential part of the definition, since the reference prior maximizes the *missing* information, which is the expected discrepancy between prior knowledge and *perfect* knowledge.
- Reference priors only depend on the asymptotic behavior of the model, which greatly simplifies their derivation. For example, in one-parameter models and under appropriate regularity conditions to guarantee asymptotic normality, the reference prior is simply Jeffreys' prior:

$$\pi(\theta) \propto i(\theta)^{1/2}, \qquad \text{where} \qquad i(\theta) = -\int_{\mathcal{X}} dx\, p(x \,|\, \theta)\, \frac{\partial^2}{\partial \theta^2} \ln p(x \,|\, \theta).$$

- Reference priors are independent of sample size.
- Reference priors are compatible with sufficient statistics and consistent under reparametrization, due to the fact that the expected information is invariant under such transformations.
- Reference priors do not represent subjective belief and should not be interpreted as prior probability distributions. In fact, they are often improper. Only reference *posteriors* have a probability interpretation.

Reference Priors in the Presence of Nuisance Parameters

Suppose the statistical model is $p(x \,|\, \theta, \lambda)$, where θ is the parameter of interest and λ is a nuisance parameter. We now need a joint reference prior $\pi(\theta, \lambda)$. The algorithm is sequential:

1. Hold θ constant and apply the one-parameter reference algorithm to obtain the conditional reference prior $\pi(\lambda \,|\, \theta)$.
2. Derive the one-parameter integrated model:
$$p(x \,|\, \theta) \;=\; \int_{\Lambda} d\lambda \; p(x \,|\, \theta, \lambda) \; \pi(\lambda \,|\, \theta).$$
3. Apply the one-parameter reference algorithm again, this time to $p(x \,|\, \theta)$, and obtain the marginal reference prior $\pi(\theta)$.
4. Set $\pi(\theta, \lambda) = \pi(\lambda \,|\, \theta) \, \pi(\theta)$.

Note that step 2 will not work if $\pi(\lambda \,|\, \theta)$ is improper ($p(x \,|\, \theta)$ will not be normalizable). The solution in that case is to introduce a sequence $\{\Lambda_i\}_{i=1}^{\infty}$ of subsets of Λ that converges to Λ and such that $\pi(\lambda \,|\, \theta)$ is integrable over each Λ_i. The integration at step 2 is then performed over Λ_i instead of Λ. This procedure results in a sequence of posteriors $\{\pi_i(\theta \,|\, x)\}_{i=1}^{\infty}$ and the desired reference posterior is obtained as the limit of that sequence.

Reference Priors: Comments

- Generalization of the reference algorithm from two to any number of parameters is straightforward.

- Since the algorithm is sequential, it requires that the parameters be ordered, say in order of inferential interest. In most applications it is found that the order does not affect the result, but there are exceptions.

- A direct consequence of this sequential algorithm is that, within a *single* model, it is possible to have as many reference priors as there are possible parameters of interest. This is because a setup that maximizes the missing information about a parameter θ will generally differ from a setup that maximizes the missing information about a parameter η, unless η is a one-to-one function of θ.

- The good news is that using different non-subjective priors for different parameters of interest is the *only* way to avoid the marginalization paradoxes.

Restricted Reference Priors

The definition of reference priors specifies that they must be taken from a class $\mathcal{P}$ of priors that are compatible with whatever initial information is available. If there is no initial information, the class is labeled $\mathcal{P}_0$ and the prior is unrestricted. Initial information can take several forms:

1. Constraints on parameter space.
2. Specified expected values.
   Suppose that the initial information about θ is of the form $\mathrm{E}[g_i(\theta)] = \beta_i$, for appropriately chosen functions $g_i, i = 1, \ldots, m$. It can then be shown that the reference prior $\pi(\theta \,|\, \mathcal{M}, \mathcal{P})$ must be of the form:
   $$\pi(\theta \,|\, \mathcal{M}, \mathcal{P}) \;=\; \pi(\theta \,|\, \mathcal{M}, \mathcal{P}_0)\; \exp\left\{ \sum_{i=1}^{m} \lambda_i \, g_i(\theta) \right\},$$
   where the λ_i's are constants determined by the constraints which define $\mathcal{P}$.
3. Subjective marginal prior.
   Suppose the model depends on two parameters, θ_1 and θ_2, and the subjective marginal $\pi(\theta_1)$ is known. The reference conditional $\pi(\theta_2 \,|\, \theta_1)$ is then proportional to $|\Sigma_{22}(\theta_1, \theta_2)|^{1/2}$, where $\Sigma_{22}(\theta_1, \theta_2)$ is the per observation Fisher information for θ_2, given that θ_1 is held fixed. If the resulting $\pi(\theta_2 \,|\, \theta_1)$ is improper, it must be corrected via a sequence of compact subsets argument in order to preserve the information based interpretation.

Example: a Poisson Process with Uncertain Mean

Consider the likelihood:

$$\mathcal{L}(\sigma, \epsilon, b \,|\, n) \;=\; \frac{(b + \epsilon\sigma)^n}{n!}\, e^{-b-\epsilon\sigma},$$

where the parameter of interest is a signal cross section σ, whereas the sensitivity ϵ and background b are nuisance parameters.

Note that σ, ϵ, and b are not identifiable from n. This problem is usually addressed by introducing a subjective prior for ϵ and b, say $\pi(\epsilon, b)$.

A common choice of prior for σ is $\pi(\sigma) = 1$ (improper!), the claim being that this is noninformative. . . Whatever one may think of this claim, if the ϵ prior has non-zero density at $\epsilon = 0$ (such as a truncated Gaussian), the posterior will be improper.

Reference Prior for the Poisson Problem

Assume we are given a subjective prior $\pi(\epsilon, b)$. We must therefore find the conditional reference prior $\pi(\sigma \,|\, \epsilon, b)$. As described before, we start by calculating the Fisher information for σ given that ϵ and b are held fixed:

$$\Sigma_{\sigma\sigma} \;=\; \mathrm{E}\left[-\frac{\partial^2}{\partial\sigma^2}\ln\mathcal{L}\right] \;=\; \frac{\epsilon^2}{b+\epsilon\sigma},$$

which would suggest:

$$\pi(\sigma \,|\, \epsilon, b) \;\propto\; \frac{\epsilon}{\sqrt{b+\epsilon\sigma}}.$$

However, this is not the reference prior for this problem. Proper application of the definition requires that the above expression be renormalized using a sequence of nested compact sets, say of the form $[0, u]$, with $u > 0$:

$$\pi_u(\sigma \,|\, \epsilon, b) \;=\; \frac{\epsilon}{\sqrt{b+\epsilon\sigma}}\;\frac{\mathbf{1}(u \geq \sigma)}{2\sqrt{b+\epsilon u} \;-\; 2\sqrt{b}}.$$

The correct conditional reference prior is then given by:

$$\pi(\sigma \,|\, \epsilon, b) \;=\; \lim_{u\to\infty}\frac{\pi_u(\sigma \,|\, \epsilon, b)}{\pi_u(\sigma_0 \,|\, \epsilon_0, b_0)} \;\propto\; \sqrt{\frac{\epsilon}{b+\epsilon\sigma}},$$

with $(\sigma_0, \epsilon_0, b_0)$ any fixed point.

Reference Posterior for the Poisson Problem

To fix ideas, consider a product of gamma densities for the prior $\pi(\epsilon, b)$:

$$\pi(\epsilon, b) \;=\; \frac{\tau(\tau\epsilon)^{x-1/2}\, e^{-\tau\epsilon}}{\Gamma(x+1/2)} \, \frac{c(cb)^{y-1/2}\, e^{-cb}}{\Gamma(y+1/2)}.$$

The σ-reference posterior is then:

$$\pi(\sigma\,|\,n) \;\propto\; \int_0^\infty d\epsilon \int_0^\infty db\, \frac{(b+\epsilon\sigma)^{n-1/2}\, e^{-b-\epsilon\sigma}}{n!} \, \frac{\sqrt{\tau}(\tau\epsilon)^{x}\, e^{-\tau\epsilon}}{\Gamma(x+1/2)} \, \frac{c(cb)^{y-1/2}\, e^{-cb}}{\Gamma(y+1/2)}.$$

The integrals may seem daunting, but there is a simple Monte Carlo algorithm to generate σ values from the posterior:

1. Generate $\epsilon \sim \mathrm{Gamma}(x, 1/\tau)$;
2. Generate $b \sim \mathrm{Gamma}(y+1/2, 1/c)$;
3. Generate $t \sim \mathrm{Gamma}(n+1/2, 1)$;
4. If $t < b$, go back to step 2;
5. Set $\sigma = (t-b)/\epsilon$;

where $\mathrm{Gamma}(z\,|\,\alpha, \beta) \equiv z^{\alpha-1} e^{-z/\beta} / \Gamma(\alpha)\beta^{\alpha}$.

Coverage of Reference Bayes Poisson Upper Limits

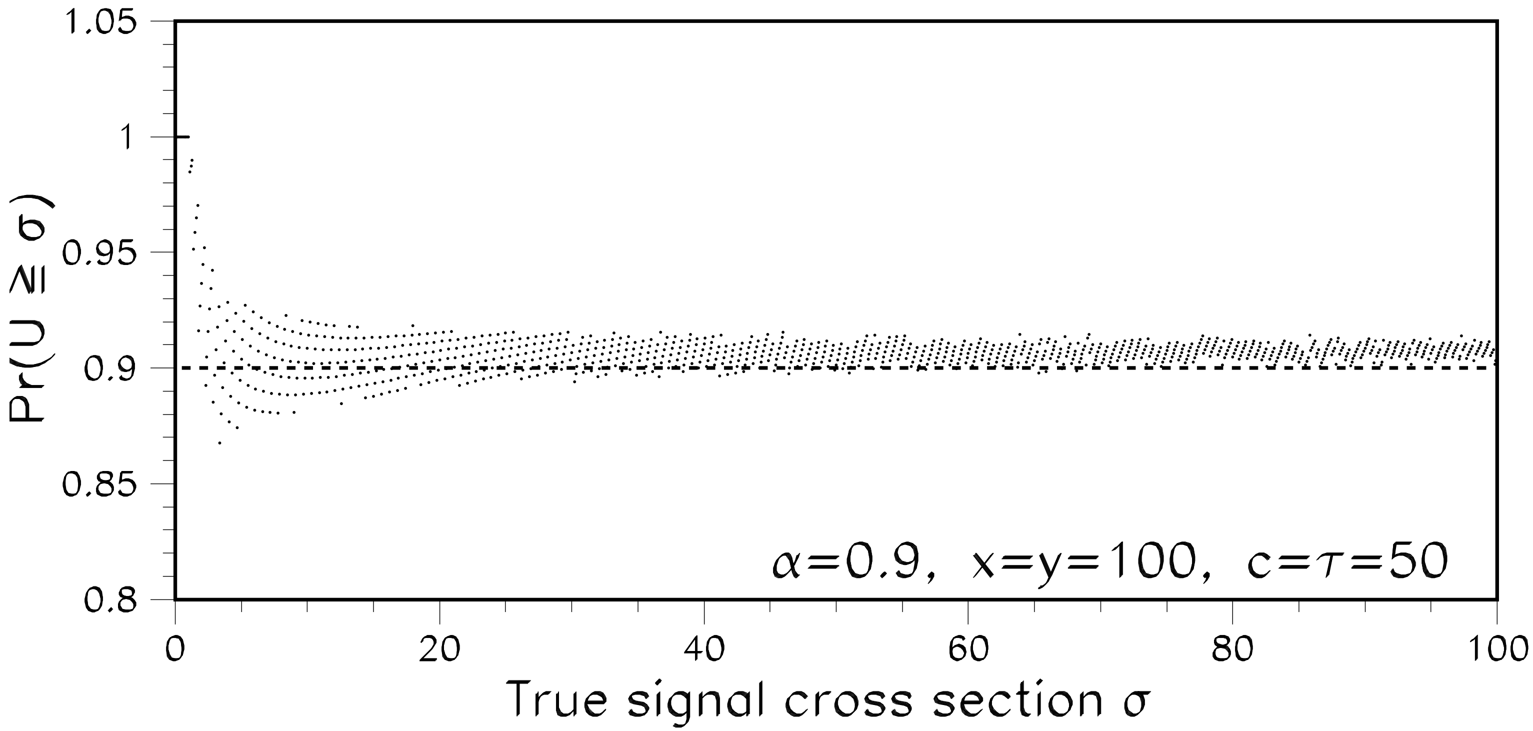

Figure 1: Coverage of 90% credibility level reference Bayes upper limits on a signal cross section σ, as a function of the true value of that cross section.

Intrinsic Estimation and Intrinsic Credible Regions (1)

The Bayesian outcome of a problem of inference is always the *full* posterior distribution for the parameter of interest.

However, it is often useful and sometimes even necessary to *summarize* the posterior distribution by providing a measure of location and quoting regions of given posterior probability content.

The typical Bayesian approach formulates point estimation as a decision problem. Suppose that $\hat{\theta}$ is an estimate of the parameter θ, whose true value θ_t is unknown. One specifies a loss function $\ell(\hat{\theta}, \theta_t)$, which measures the consequence of using the model $p(x \,|\, \hat{\theta})$ instead of the true model $p(x \,|\, \theta_t)$. The Bayes estimator $\theta_b = \theta_b(x)$ of the parameter θ minimizes the corresponding posterior loss:

$$\theta_b(x) \;=\; \arg \min_{\hat{\theta} \in \Theta} \int_{\Theta} d\theta \; \ell(\hat{\theta}, \theta) \; p(\theta \,|\, x).$$

Some conventional loss functions are:

1. <u>*Squared error loss:*</u> $\ell(\hat{\theta}, \theta_t) = (\hat{\theta} - \theta_t)^2 \quad \Rightarrow \quad \theta_b$ is the *posterior mean*.
2. <u>*Zero-one loss:*</u> $\ell(\hat{\theta}, \theta_t) = 1 - \mathrm{I}_{[\theta_t - \epsilon, \theta_t + \epsilon]}(\hat{\theta}) \quad \Rightarrow \quad \theta_b$ is the *posterior mode*.
3. <u>*Absolute error loss:*</u> $\ell(\hat{\theta}, \theta_t) = |\hat{\theta} - \theta_t| \quad \Rightarrow \quad \theta_b$ is the *posterior median*.

Intrinsic Estimation and Intrinsic Credible Regions (2)

In physics, interest usually focuses on the actual mechanism that governs the data. Therefore we need a point estimate that is invariant under one-to-one transformations of the parameter and/or the data (including reduction to sufficient statistics). Fortunately, we have already encountered a loss function that will deliver such an estimate: the intrinsic discrepancy!

The intrinsic discrepancy between two probability densities p_1 and p_2 is:

$$\delta\{p_1, p_2\} = \min\left\{\int dx\, p_1(x)\, \ln\frac{p_1(x)}{p_2(x)},\ \int dx\, p_2(x)\, \ln\frac{p_2(x)}{p_1(x)}\right\},$$

provided one of the integrals is finite. The intrinsic discrepancy between two parametric models for x,
$\mathcal{M}_1 = \{p_1(x\,|\,\phi), x \in \mathcal{X}, \phi \in \Phi\}$ and $\mathcal{M}_2 = \{p_2(x\,|\,\psi), x \in \mathcal{X}, \psi \in \Psi\}$,
is the minimum intrinsic discrepancy between their elements:

$$\delta\{\mathcal{M}_1, \mathcal{M}_2\} = \inf_{\phi,\psi} \delta\{p_1(x\,|\,\phi), p_2(x\,|\,\psi)\}.$$

This suggests setting $\ell(\hat{\theta}, \theta_t) = \delta\{\hat{\theta}, \theta_t\} \equiv \delta\{p(x\,|\,\hat{\theta}), p(x\,|\,\theta_t)\}$.

Intrinsic Estimation and Intrinsic Credible Regions (3)

Let $\{p(x\,|\,\theta), x \in \mathcal{X}, \theta \in \Theta\}$ be a family of probability models for some observable data x. The intrinsic estimator minimizes the reference posterior expectation of the intrinsic discrepancy:

$$\theta^\star(x) \;=\; \arg\min_{\hat\theta\in\Theta}\, d(\hat\theta\,|\,x) \;=\; \arg\min_{\hat\theta\in\Theta} \int_\Theta d\theta\,\delta\{\hat\theta,\theta\}\,\pi_\delta(\theta\,|\,x),$$

where $\pi_\delta(\theta\,|\,x)$ is the reference posterior when the intrinsic discrepancy is the parameter of interest.

An intrinsic α-credible region is a subset $R^\star_\alpha$ of the parameter space Θ such that:

$$(i) \qquad \int_{R^\star_\alpha} d\theta\,\pi(\theta\,|\,x) \;=\; \alpha;$$

$$(ii) \qquad \text{For all } \theta_i \in R^\star_\alpha \text{ and } \theta_j \notin R^\star_\alpha, \quad d(\theta_i\,|\,x) \le d(\theta_j\,|\,x).$$

Although the concepts of intrinsic estimator and credible region have been defined here for *reference* problems, they can also be used in situations where proper prior information is available.

Example: Transverse Momentum Measurement (1)

Consider the measurement of the transverse momentum of particles in a tracking chamber immersed in a magnetic field. The probability density is (approximately) Gaussian in the inverse of the transverse momentum:

$$p(x\,|\,\mu) \;=\; \frac{e^{-\frac{1}{2}\left(\frac{1/x-1/\mu}{\sigma}\right)^2}}{\sqrt{2\pi}\,\sigma\,x^2},$$

where x is the measured signed p_T, μ is the true signed p_T, and σ is a function of the magnetic field strength and the chamber resolution.

It is easy to verify that a naive Bayesian analysis yields unreasonable results. To begin with, "non-informative" priors such as $\pi(\mu) \propto 1$ or $\pi(\mu) \propto 1/\mu$ lead to improper posteriors. The next choice, $\pi(\mu) \propto 1/\mu^2$, does lead to a proper posterior, but the resulting HPD Bayes estimate of μ is bounded from above, regardless of the measured value x! Similarly, HPD intervals always exclude μ values above a certain threshold, with the consequence that their coverage drops to zero above that threshold.

One would think that a reference analysis of this problem will yield a more satisfactory solution due to its invariance properties.

Example: Transverse Momentum Measurement (2)

Fortunately, a reference analysis of this problem can be done entirely analytically:

1. Intrinsic discrepancy:

$$\delta\{\hat{\mu},\mu\} \;=\; \frac{1}{2}\left(\frac{1/\mu - 1/\hat{\mu}}{\sigma}\right)^2.$$

2. Reference prior when μ is the quantity of interest: $\pi(\mu) \propto 1/\mu^2$.
3. Reference prior when δ is the quantity of interest. Since δ is a piecewise one-to-one function of μ, this reference prior is also $1/\mu^2$.
4. Reference posterior:

$$p(\mu\,|\,x) \;=\; \frac{e^{-\frac{1}{2}\left(\frac{1/x-1/\mu}{\sigma}\right)^2}}{\sqrt{2\pi}\,\sigma\,\mu^2}.$$

5. Reference posterior expected intrinsic loss:

$$d(\hat{\mu}\,|\,x) \;=\; \frac{1}{2} + \frac{1}{2}\left(\frac{1/x - 1/\hat{\mu}}{\sigma}\right)^2.$$

Example: Transverse Momentum Measurement (3)

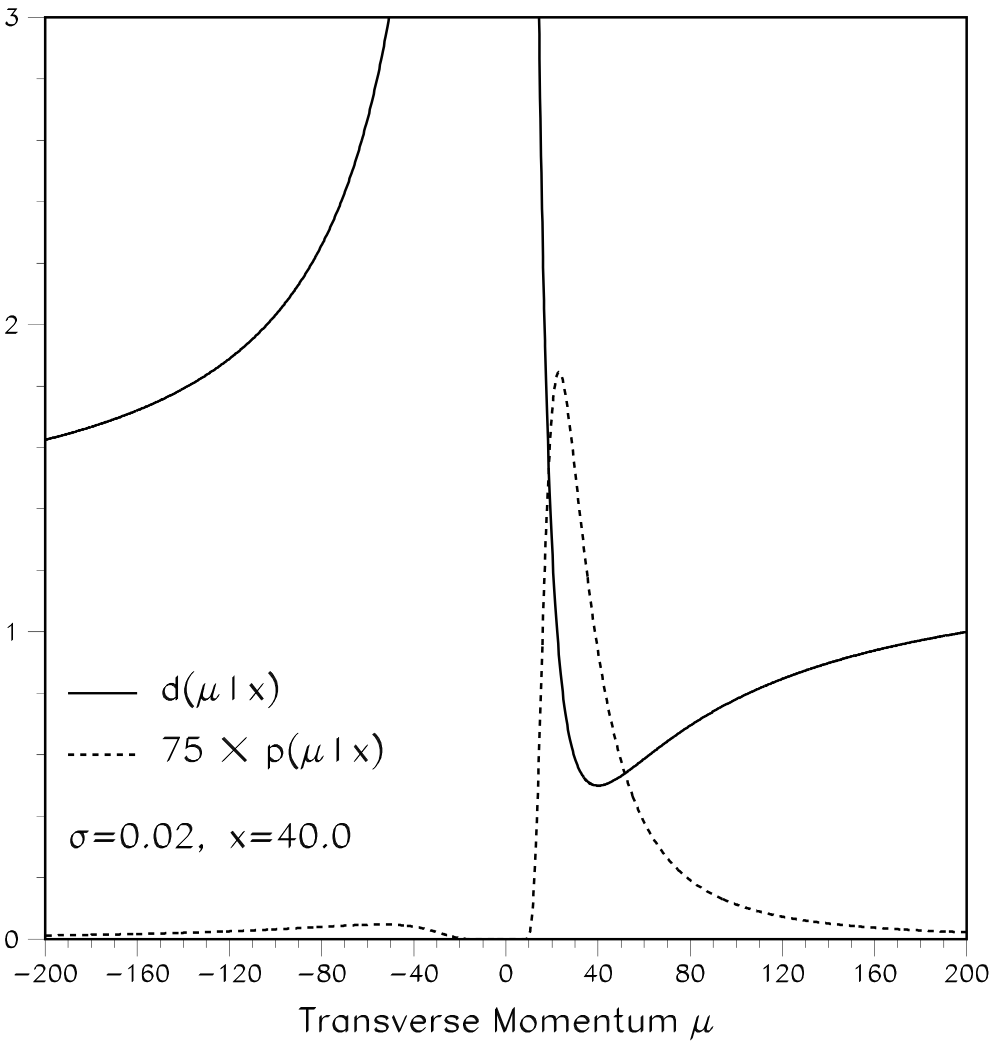

Figure 2: Reference posterior expected intrinsic loss $d(\mu \,|\, x)$ (solid line), and reference posterior density $p(\mu \,|\, x)$ (dashed line) for the problem of measuring transverse momenta in a tracking chamber.

Example: Transverse Momentum Measurement (4)

The results of the reference analysis are as follows:

- The intrinsic estimate of μ, i.e. the value of μ that minimizes the reference posterior expected intrinsic loss, is $\mu^\star = x$.

- Minimum reference posterior expected intrinsic loss intervals have the form:

$$\text{If } d < \frac{1}{2} + \frac{1}{2\sigma^2 x^2}: \qquad \left[\frac{x}{1+\sigma x\sqrt{2d-1}}, \frac{x}{1-\sigma x\sqrt{2d-1}}\right],$$

$$\text{If } d = \frac{1}{2} + \frac{1}{2\sigma^2 x^2} \text{ and } x \geq 0: \qquad \left[\frac{x}{2}, +\infty\right],$$

$$\text{If } d = \frac{1}{2} + \frac{1}{2\sigma^2 x^2} \text{ and } x < 0: \qquad \left[-\infty, \frac{x}{2}\right],$$

$$\text{If } d > \frac{1}{2} + \frac{1}{2\sigma^2 x^2}: \qquad \left[-\infty, \tfrac{x}{1-\sigma x\sqrt{2d-1}}\right] \cup \left[\tfrac{x}{1+\sigma x\sqrt{2d-1}}, +\infty\right],$$

  where d is determined by the requirement of a specified posterior probability content. Note that $\mu^\star$ is contained in all the intrinsic intervals.

Summary

- Noninformative priors have been studied for a long time and most of them have been found defective in more than one way. Reference analysis arose from this study as the only *general* method that produces priors that have the required *invariance* properties, deal successfully with the *marginalization* paradoxes, and have consistent *sampling* properties.

- Reference priors should not be interpreted as probability distributions expressing subjective degree of belief; instead, they help answer the question of what could be said about the quantity of interest if one's prior knowledge were dominated by the data.

- Reference analysis also provides methods for summarizing the posterior density of a measurement. Intrinsic point estimates, credible intervals, and hypothesis tests have invariance properties that are essential for *scientific* inference.

- Much more information can be found on José Bernardo's web page, `http://www.uv.es/~bernardo/publications.html`.