# Probability Matching Priors in LHC Physics

*P.D. Baines and X.-L. Meng*
Department of Statistics, Harvard University

**Abstract**
Probability matching priors (PMPs) provide a bridge between Bayesian and frequentist inference by yielding Bayesian posterior intervals with frequentist validity. PMPs are, in general, challenging to implement as they are defined as solutions to a potentially high-dimensional and non-linear PDE. Outside the orthogonal case, no general framework exists for the implementation of PMPs. Recent work has made progress in this area, although no approach can yet be applied in generality. We consider PMPs for the three Poisson system arising in LHC experiments. Connections to reference and reverse reference priors are also considered. Theoretical and simulation results are presented, with comparison to other Bayesian techniques.

## 1 The Problem & Motivation

The problem of reliably estimating the intensity of a ‘signal’ in the presence of background and calibration uncertainties is a common one in LHC Physics and throughout the scientific world. Here we consider application of a class of Bayesian prior distributions to this problem, known as *probability matching priors* (PMPs). PMPs provide a bridge between the two main paradigms of statistical inference: frequentist and Bayes. Direct implementation of PMPs is, in general, extremely challenging as a result of possibly high-dimensional and non-linear partial differential equations (PDEs) that must be solved. This paper introduces both the rich rewards that may be reaped from applying PMPs in LHC Physics analyses, as well as the challenges that must first be overcome.

The primary criterion for the methods considered here will be their coverage properties. Other criteria such as credibility, length, bias and behavior in ‘boundary’ cases are also of importance and shall be addressed where space permits. PMPs are constructed to have (approximate) frequentist validity, will have good credibility over a range of prior distributions, and avoid many undesirable properties such as zero-length intervals. In this sense, where the desired coverage can be achieved, PMPs would appear to provide an ‘optimal’ solution likely to be accessible to both Bayesians and frequentists. However, existence of a PMP is not guaranteed. In the LHC example presented here, a large class of candidate priors are shown to not be PMPs.

While the theoretical properties of PMPs are well understood (see Ref. [1] for a review), their implementation remains an immense challenge. Recent papers by Levine & Casella [2] and Sweeting [3] have attempted to address this challenge, albeit not yet in full generality. In section 2 we provide a brief introduction to PMPs and orthogonality. Implementation is discussed in section 3, with an LHC application presented in section 4. Brief discussion is provided in section 5.

## 2 Introduction to Probability Matching Priors

### 2.1 Probability Matching Priors

The definition of a PMP for $\psi \in \mathbb{R}$, is that the posterior quantiles of $\psi$ have (approximate) frequentist validity. See Ref. [1] for a formal definition. Peers [4] derived a PDE that a prior distribution must satisfy if it is to be first order probability matching (PM) (i.e., coverage of $\psi^{(1-\alpha)}$, the $100(1-\alpha)$ posterior percentile of $\psi$, is $1-\alpha+o(n^{-1/2})$ for all $0<\alpha<1$, where $n$ is the sample size).

**Theorem 1** ***First Order PMP Condition**: Let $\psi$ be a univariate parameter of interest, with $\phi \in \mathbb{R}^{p-1}$ a nuisance parameter. The data are assumed to be generated from the family $f(\cdot;\psi,\phi)$. Let $I_{ij}$ and $I^{ij}$ denote the corresponding elements of the Fisher Information matrix and its inverse respectively. A prior $\pi(\cdot)$ is first order PM if and only if it satisfies the PDE:*

$$\frac{\partial}{\partial\psi}\left\{\pi(\psi,\phi)\cdot(I^{\psi\psi})^{1/2}\right\}+\sum_{j=1}^{p-1}\frac{\partial}{\partial\phi_j}\left\{\pi(\psi,\phi)I^{\phi_j\psi}(I^{\psi\psi})^{-1/2}\right\}=0. \tag{1}$$

Analytic solutions to this generally nonlinear $p-$dimensional PDE are rarely possible, and numerical solutions are often equally as elusive. However, in the case of an orthogonal parameterisation, that is, $I^{\psi,\phi_j} = 0$ for all $j$, the solution is trivially given by:

$$\pi(\psi,\phi) = I_{\psi\psi}^{1/2}\cdot d(\phi) \tag{2}$$

where $d(\phi)$ is an arbitrary smooth function of the nuisance parameter (see Tibshirani, Ref. [5]). We, therefore, naturally attempt to extend the utility of (2) even when the parameterisation fails to be exactly orthogonal. The arbitrary function $d(\phi)$ can have a strong impact on finite-sample properties: the reverse reference prior [6] is a recommended tool for selecting within this class.

### 2.2 Orthogonality

The formal definition of orthogonality, from Cox & Reid [7], is that the partitioned Fisher Information (FI) is block diagonal, that is, $I^{\psi,\phi_j} = 0$ for all $j$. Cox & Reid showed that for a scalar parameter of interest there always exists a transformation to achieve orthogonality with a $(p-1)-$dimensional nuisance parameter. However, the transformation is defined as the solution to a set of $(p-1)$ PDE's. These equations are in general not solvable by standard methods, and pose arguably a greater challenge than the PMP PDE (1). Hence, two obvious routes to finding probability matching priors, from the definition and via orthogonal parameterisation, are blocked by the obstacle of an intractable (set of) PDE(s). A third route is to derive either the reference prior of Berger and Bernardo [8], or reverse reference prior and check whether it is probability matching (frequently they are). However, outside the orthogonal case, their derivation can also become extremely challenging.

## 3 Existing Implementation Methods & Their Limitations

Levine & Casella [2] (LC) describe a Monte Carlo scheme to sample from the posterior distribution under a prior that is a solution to (1), when the nuisance parameter is univariate. The high run-time for the algorithm also makes it infeasible for large-scale simulation studies as considered here. Sweeting [3] proposes a more general approach that removes the restriction to univariate nuisance parameters, by seeking a *local probability matching prior*, using data-dependent approximations. The approach requires a non-trivial condition on the parameterisation, a condition that is not satisfied in the LHC application of section 4. In the general case it is unclear how to construct a parameterisation satisfying the condition if one is not immediately obvious. Indeed, in the LHC examples of section 4 the condition is not satisfied.

## 4 LHC Physics Example

The following problem is a common one in LHC Physics. The parameter of interest, $s$, represents the signal, monitored for $M$ decay channels, with $\epsilon_i$ and $b_i$ unknown channel-specific effective area and background parameters. Consider,

$$n_i|s,\epsilon_i,b_i \sim Pois\left(\epsilon_i s + b_i\right), \qquad y_i|b_i \sim Pois\left(t_i b_i\right), \qquad z_i|\epsilon_i \sim Pois\left(u_i\epsilon_i\right), \tag{3}$$

with $i = 1,\ldots,M$, $\{t_1,\ldots,t_M,u_1,\ldots,u_M\}$ known constants and observations assumed to be independent. The goal is to find a PMP for $s$ under this model. For simplicity we consider only the single channel

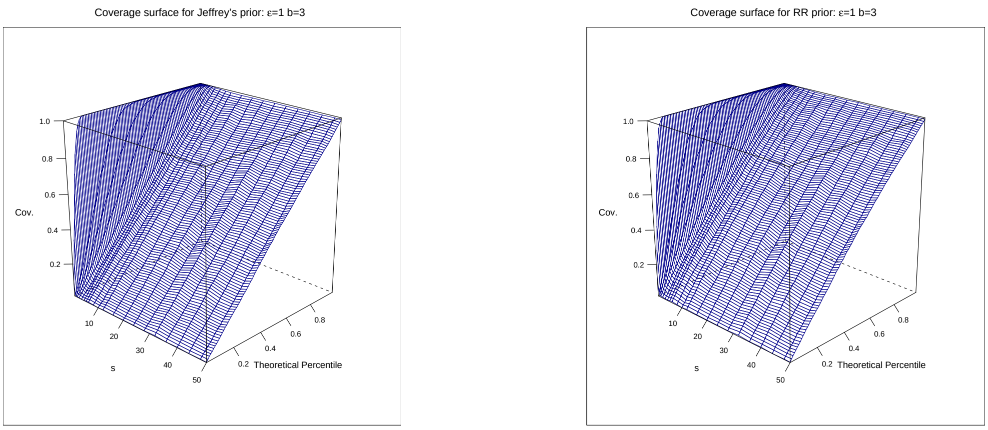

**Fig. 1:** Coverage surfaces for Jeffreys' [L] and the reverse reference [R] priors. The $z-$axis displays the coverage, $x$ and $y$-axes indicate nominal coverage and the value of $s$. A 'perfect coverage method' would give a plane at 45°.

($M = 1$) case and drop the subscripts. The multi-channel setting is known to be more challenging, see Heinrich [9]. The first order PMP PDE can be shown to be:

$$\frac{\partial}{\partial s}\left\{\pi\sqrt{\frac{\epsilon st(u+s)+bu(1+t)}{\epsilon^2 tu}}\right\}-\frac{\partial}{\partial b}\left\{\pi\sqrt{\frac{b^2u}{\epsilon st^2(u+s)+but(1+t)}}\right\}-\frac{\partial}{\partial \epsilon}\left\{\pi\sqrt{\frac{s^2\epsilon t}{\epsilon stu(u+s)+bu^2(1+t)}}\right\}=0 \tag{4}$$

This cannot be directly solved by standard software, which may suggest that no solution exists. The prior from (2) here becomes $\pi(s,b,\epsilon)\propto d(b,\epsilon)/\sqrt{s\epsilon+b}$. Jeffreys prior is found to be a special case, where $d(b,\epsilon)=\sqrt{\epsilon tu/b}$. This is not the case for $M>1$. In all cases posterior propriety must be checked. The general $M-$channel reverse reference prior $\pi_{rr}$ can be fully derived. The regular reference prior $\pi_r$ for the ordered parameterisation $\psi = s, \phi = (\mathbf{b},\epsilon)$, if it exists, is of the form:

$$\pi_{rr}(s,\mathbf{b},\epsilon) \quad \propto \quad \sqrt{\frac{\sum_{j=1}^M \epsilon_j u_j}{\prod_{j=1}^M b_j\epsilon_j}\cdot\sum_{j=1}^M\frac{\epsilon_j^2}{s\epsilon_j+b_j}\cdot\frac{1}{\sum_{j=1}^M\epsilon_j}}\,, \tag{5}$$

$$\pi_r(s,\mathbf{b},\epsilon) \quad \propto \quad g(s)\sqrt{\prod_{j=1}^M\frac{b_ju_j(1+t_j)+\epsilon_j st_j(s+u_j)}{b_j\epsilon_j(b_j+\epsilon_j s)}}\,. \tag{6}$$

By plugging in the form of the prior distribution into (4), it can be proved that, for the single-channel case, neither the regular reference prior nor any priors within the Tibshirani class of priors from (2) can be a PMP. For example, plugging the reference prior into (1), we obtain an ODE for the function $g(s)$. However, this ODE can be shown to have no solution. An analogous proof holds for the Tibshirani class from (2), hence, also the reverse reference prior. These results, combined with the failure to directly solve (4), strongly suggest that there in fact may be no PMP in this example.

Instead, we considered three priors of the form (2):

$$d_J(b,\epsilon)=\sqrt{\epsilon/b} \qquad d(b,\epsilon)=1/\sqrt{b\epsilon} \qquad d(b,\epsilon)=1, \tag{7}$$

where $d_J$ corresponds to Jeffreys prior and $d = 1/\sqrt{b\epsilon}$ to a form of pseudo-Jeffreys' prior for $b$ and $\epsilon$. For comparison, priors of the form $\pi(s,b,\epsilon)\propto\frac{1}{\sqrt{s}}$ and $\frac{1}{s}$ are also considered. 110,000 datasets were simulated from (3) with $b = 3$, $\epsilon = 1$, $t = 33.0$, $u = 100.0$ and with $s$ taking on 22 values in the range 0.1 to 48.0. Posterior intervals were obtained under all of the above prior distributions. Figure 1 displays the coverage surface for Jeffreys' prior. Numerical results are presented in Table 1. Both Jeffreys' and the $d = \frac{1}{\sqrt{b\epsilon}}$ prior have excellent coverage properties over a wide range of $s$. For $M \geq 8$, say, coverage properties often deteriorate. Overcoverage for small $s$ is inevitable under the Bayesian methodology, and a necessary price to pay for any method that does not produce zero-length intervals.

**Table 1:** For $s = 20$, the actual coverage of nominal $5, 10, 25, 50, 75, 90, 95$ & $99^{\text{th}}$ percentiles produced by using each of the different priors discussed in section 4

| $s^{(\alpha)}$ | $\pi_{rr}$: $d = \frac{1}{\sqrt{b}}$ | Jeffreys': $d = \sqrt{\frac{\epsilon}{b}}$ | $d = \frac{1}{\sqrt{b\epsilon}}$ | $d = 1$ | $\pi \propto 1$ | $\pi \propto \frac{1}{\sqrt{s}}$ |
|---|---|---|---|---|---|---|
| $s^{(0.05)}$ | 0.06 | 0.05 | 0.05 | 0.05 | 0.06 | 0.06 |
| $s^{(0.10)}$ | 0.12 | 0.10 | 0.11 | 0.11 | 0.12 | 0.12 |
| $s^{(0.25)}$ | 0.29 | 0.25 | 0.27 | 0.28 | 0.29 | 0.29 |
| $s^{(0.50)}$ | 0.54 | 0.49 | 0.51 | 0.52 | 0.54 | 0.54 |
| $s^{(0.75)}$ | 0.78 | 0.74 | 0.75 | 0.76 | 0.78 | 0.78 |
| $s^{(0.90)}$ | 0.91 | 0.89 | 0.90 | 0.91 | 0.91 | 0.91 |
| $s^{(0.95)}$ | 0.96 | 0.95 | 0.95 | 0.96 | 0.96 | 0.96 |
| $s^{(0.99)}$ | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |

## 5 Discussion

Since the primary goal is to produce intervals with frequentist validity: “why not just be a frequentist?” One benefit of PMPs is that they produce intervals accessible to frequentists and Bayesians alike. Moreover, this accessibility is independent of the criteria by which they are evaluated. Other criteria are discussed in Heinrich [9]. In many of these respects PMPs may provide a more satisfactory solution than other methods produced from frequentist principles alone. For example, as discussed in Heinrich [9], both frequentist and likelihood-based methods can produce undesirable zero-length intervals. This behaviour cannot occur under the Bayesian construction presented here, a side-effect of this is overcoverage for small signal $s$.

PMPs, where they exist, may provide an ‘optimal’ solution to coverage problems. In the LHC example considered here, no exact PMP has been found so far, but approximate PMPs seem to exist over restricted ranges of the parameter space, and may be all that is required for practical purposes. Reference and reverse priors are also recommended as an effective default prior for Bayesian inference, often satisfying the PMP property. Further progress on both computational issues and operational properties will help give practitioners another option for making reliable inference about important physical parameters arising in LHC experiments.

## References

[1] Datta, G.S. & Mukerjee, R., (2004) Probability Matching Priors: Higher Order Asymptotics. *Lecture Notes in Statistics 178, Springer-Verlag.*

[2] Levine, R.A. & Casella, G. (2003) Implementing probability matching priors for frequentist inference. *Biometrika* **90**, 127-137.

[3] Sweeting, T.J. (2005) On the implementation of local probability matching priors for interest parameters. *Biometrika* **92**, 47-57.

[4] Peers, H.W. (1965) On Confidence sets and Bayesian probability points in the case of several parameters. *J. of the Royal Stat. Soc. B* **53**, 611-618.

[5] Tibshirani, R.J. (1989) Noninformative priors for one parameter of many. *Biometrika* **76**, 604-608.

[6] Berger, J.O. (1993) Discussion of “Non-informative Priors” by Ghosh, J.K, Mukerjee, R. *Bayesian Statistics 4*, 205-206

[7] Cox, D.R. & Reid, N. (1987) Parameter orthogonality and approximate conditional inference (with discussion). *J. of the Royal Stat. Soc. B* **49**, 1-39.

[8] Berger, J.O, & Bernardo, J.M. (1992) On the Development of Reference Priors. *Bayesian Statistics 4*, 35-60.

[9] Heinrich, J. (2007) Report on the Banff Limits Challenge.
`http://newton.hep.upenn.edu/~heinrich/challenge.pdf`