DIP – Diagnostics for Insufficiencies of Posterior calculations in Bayesian signal inference

Sebastian Dorn*[,1,2], Niels Oppermann[1], Torsten A. Enßlin[1]
[1] *Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany*
[2] *Technische Universität München, Arcisstraße 21, 80333 München, Germany*

(Dated: July 16, 2013)

Abstract. We present an error-diagnostic validation method for posterior distributions in Bayesian signal inference. It transfers deviations from the correct posterior into characteristic deviations from a uniform distribution of a quantity constructed for this purpose. We show that this method is able to reveal and discriminate several kinds of numerical and approximation errors. For this we present a number of analytical examples of posteriors with incorrect variance, skewness, position of the maximum, or normalization. We show further how this test can be applied to multidimensional signals.

Subject headings: critical test - Bayesian inference - software validation - error diagnostics.

Introduction. Bayesian inference methods are gaining importance in many areas of physics, like e.g. precision cosmology [1]. Dealing with Bayesian models means to grapple with the posterior probability distribution, whose calculation and simulation is often highly complex and therefore prone to errors. Rather than taking the correctness of the numerical implementation of the posterior for granted, one should validate it in some way.

Although there are validation approaches (e.g. [2, 3]), these provide little diagnostics for the type of error. However, this information would be very useful in order to locate a mistake in a posterior calculating code or in its mathematical derivation. Therefore we introduce a validation method that is able to detect errors in the numerical implementation as well as in the mathematical derivation. The typical deviation of a quantity constructed for this purpose from a uniform distribution encodes information on the kind of errors made.

Posterior validation method. Within this work we assume a data set d is given in the form $d = (d_1, d_2, \ldots, d_m)^T \in \mathbb{R}^m$, where $m \in \mathbb{N}$, and we want to extract a physical quantity, $s \in \mathbb{R}$, from its posterior probability density function (PDF), $P(s|d)$. The data is drawn from the likelihood $P(d|s)$,

$$d \hookleftarrow P(d|s). \qquad (1)$$

The posterior is given by Bayes' Theorem [4],

$$P(s|d) = \frac{P(d,s)}{P(d)} = \frac{P(d|s)P(s)}{P(d)}, \qquad (2)$$

where the prior is denoted by $P(s)$ and the evidence by $P(d)$. A concrete example of such a calculation including approximations that require validation can be found in [5].

Now we introduce the *Diagnostics for Insufficiencies of Posterior calculations* (DIP). This is a validation method for the numerical calculation of the posterior $P(s|d)$. For this purpose we use the following procedure [2]:

1. Sample s_{gen} from the prior $P(s)$.
2. Generate data d for s_{gen} according to $P(d|s_{\mathrm{gen}})$.
3. Calculate a posterior curve for given data by determining $\tilde{P}(s|d)$ according to Eq. (2), where $\tilde{P}$ denotes the posterior including possible approximations.
4. Calculate the posterior probability for $s \leq s_{\mathrm{gen}}$ according to

$$x := \int_{-\infty}^{s_{\mathrm{gen}}} \mathrm{d}s\ \tilde{P}(s|d)\ \in [0,1]. \qquad (3)$$

5. If the calculation of the posterior was correct, the distribution for x, $P(x)$, should be uniform between 0 and 1.

The uniformity of $P(x)$ can then be checked numerically by going through steps 1-4 repeatedly. Note that the distribution of x can be uniform even if there is an error in the implementation or mathematical derivation. The reason for this is the unlikely possibility of at least two errors compensating each other exactly. However, this is a fundamental problem of nearly every numerical validation method.

Proof: We show here analytically that $P(x) = 1$ if $\tilde{P}(s|d) = P(s|d)$, as an alternative to the discussion in [2]:

*sdorn@mpa-garching.mpg.de

$$\begin{aligned} P(x) &= \int_{-\infty}^{\infty} \mathrm{d}s \int \mathcal{D}d \; P(x,d,s) \\ &= \int_{-\infty}^{\infty} \mathrm{d}s \int \mathcal{D}d \; P(x|d,s)P(d,s) \\ &= \int_{-\infty}^{\infty} \mathrm{d}s \int \mathcal{D}d \; P(d,s) \; \delta \left(x - \int_{-\infty}^{s} \mathrm{d}s' \; P(s'|d) \right) \\ &= \int_{-\infty}^{\infty} \mathrm{d}s \int \mathcal{D}d \; P(d)P(s|d) \; \delta(x - x_d(s)), \end{aligned} \tag{4}$$

where $x_d(s) := \int_{-\infty}^{s} \mathrm{d}s' \; P(s'|d)$ and $\int \mathcal{D}d$ denotes a path-integral over all possible realizations of d. Now we show $P(x) = 1$ for $x \in [0,1]$:

$$\begin{aligned} P(x) &= \partial_x \int_0^x \mathrm{d}x' P(x') \\ &= \partial_x \int \mathcal{D}d \; P(d) \int_{-\infty}^{\infty} \mathrm{d}s \; P(s|d) \underbrace{\int_0^x \mathrm{d}x' \; \delta(x' - x_d(s))}_{\Theta(x - x_d(s))} \\ &= \partial_x \int \mathcal{D}d \; P(d) \int_{-\infty}^{s_d(x)} \mathrm{d}s \; P(s|d) \\ &= \partial_x \int \mathcal{D}d \; P(d) \underbrace{x_d(s_d(x))}_{=x} = \partial_x x \int \mathcal{D}d \; P(d) \\ &= \partial_x x = 1 \end{aligned} \tag{5}$$

Here $s_d(x)$ is the inverse of $x_d(s)$ and Θ the Heaviside step function. This inverse exists because $x_d(s)$ is strictly monotonous, unless $P(s|d) = 0$ exactly for some s-range. □

Analytic examples of insufficient posteriors. To investigate the influence of an insufficient posterior on the distribution $P(x)$ we study as an example a Gaussian posterior,

$$P(s|d) = \mathcal{G}(s_d, \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{s_d^2}{2\sigma^2} \right), \tag{6}$$

with $s_d = s - \bar{s}_d$ and $\bar{s}_d$ the data dependent maximum of the posterior. In the following we assume the variance to be data independent and consider a wrongly determined value $x^\epsilon = \int_{-\infty}^{s_{\mathrm{gen}}} \mathrm{d}s \; P^\epsilon(s|d)$, where $P^\epsilon(s|d)$ is Gaussian with wrong variance or non-zero skewness or wrong maximum position or wrong normalization.

Wrong variance. In the case in which $P(x)$ was calculated from a posterior whose standart deviation deviates by a fraction ϵ from the true value of σ, we consider

$$\begin{aligned} P^\epsilon(s|d) &= \frac{1}{\sqrt{2\pi}\sigma(1+\epsilon)} \exp\left(-\frac{s_d^2}{2\sigma^2(1+\epsilon)^2} \right), \\ x^\epsilon &= \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{s_d}{\sqrt{2}\sigma(1+\epsilon)} \right) \right] \end{aligned} \tag{7}$$

with $\epsilon > -1$. To determine the distribution $P(x)$ we use Eq. (4). This yields

$$P(x) = (1+\epsilon) \exp\left(-\left[\mathrm{erf}^{-1}(2x-1) \right]^2 \left[(1+\epsilon)^2 - 1 \right] \right) \tag{8}$$

with the limit $P(x) \overset{\epsilon \to 0}{\longrightarrow} 1$. The deviations from the uniform distribution increase with the value of $|\epsilon|$ and are shown in Fig. 1. In case the standard deviation was underestimated, $\epsilon < 0$, the distribution for x becomes convex ("∪-shape") and for an overestimation, $\epsilon > 0$, it becomes concave ("∩-shape"). Since the underestimation of variances is a typical mistake, the DIP-test produces often test distributions with a dip in the middle.

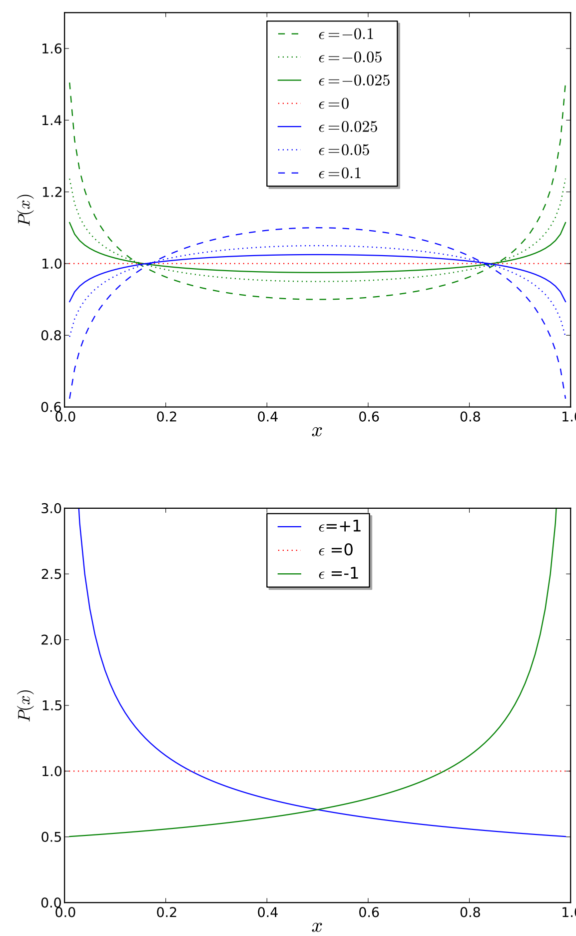

FIG. 1: Influence of an insufficient posterior on the DIP-distribution $P(x)$. The upper (lower) panel shows the effect of calculating $P(x)$ from a posterior with wrong variance (skewness) as described by Eq. (7) (Eq. (9)).

Wrong skewness. Next, we consider the case in which $P(x)$ was calculated from a falsely skewed posterior,

$$P^\epsilon(s|d) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{s_d^2}{2\sigma^2} \right) \left(1 + \mathrm{erf}\left(\frac{\epsilon s_d}{\sqrt{2}\sigma} \right) \right). \tag{9}$$

Thus, x^ϵ is given by

$$\begin{aligned} x^\epsilon =& \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{s_d}{\sqrt{2}\sigma}\right)\right] \\ &- \frac{1}{\pi}\int_0^\epsilon \mathrm{d}\tilde{\epsilon}\,\frac{\exp\left(-\frac{1}{2}\left(\frac{s_d}{\sigma}\right)^2\left(1+\tilde{\epsilon}^2\right)\right)}{1+\tilde{\epsilon}^2} \\ =:& \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{s_d}{\sqrt{2}\sigma}\right)\right] - 2T\left(\frac{s_d}{\sigma},\epsilon\right), \end{aligned} \tag{10}$$

where $T\left(\frac{s_d}{\sigma},\epsilon\right)$ is the *Owen's function* [6], and ϵ denotes the dimensionless skewness-parameter. Now we focus on $|\epsilon| = 1$ for simplicity, for which $2T\left(\frac{s_d}{\sigma},\pm1\right) = \pm\left(1-\mathrm{erf}^2\left(\frac{s_d}{\sqrt{2}\sigma}\right)\right)/4$. Applying Eq. (4) yields

$$P(x) = \begin{cases} 1/\left(2\sqrt{x}\right) & \text{if } \epsilon = 1 \\ 1/\left(2\sqrt{1-x}\right) & \text{if } \epsilon = -1 \end{cases}. \tag{11}$$

The effect of an incorrectly skewed posterior is an enhancement of values close to $x = 0$ or $x = 1$ (Fig. 1) and means that the 68% confidence interval around the expectation value (maximum of the Gaussian PDF) is falsely calculated to be asymmetric. Here, we restricted ourselves to the cases $\epsilon = \pm1$ due to their analytic treatability. Smaller deviations with $|\epsilon| < 1$ will lead to qualitatively similar but less pronounced distortions of the sampled distribution $P(x)$.

Wrong maximum position. In the case in which $P(x)$ was calculated from a posterior whose maximum has a wrong position, we consider

$$\begin{aligned} P^\epsilon(s|d) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(s_d-\epsilon)^2}{2\sigma^2}\right), \\ x^\epsilon &= \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{s_d-\epsilon}{\sqrt{2}\sigma}\right)\right]. \end{aligned} \tag{12}$$

Applying again Eq. (4) yields

$$P(x) = \exp\left(-\frac{1}{2}\left(\frac{\epsilon}{\sigma}\right)^2 - \sqrt{2}\left(\frac{\epsilon}{\sigma}\right)\ \mathrm{erf}^{-1}(2x-1)\right) \tag{13}$$

with the limit $P(x) \xrightarrow{\epsilon\to0} 1$. The resulting distribution of x for $\sigma = 1$ is shown in Fig. 2. Here, the x abundances near to $x = 0$ or $x = 1$ are enhanced, similarly to the case of incorrect skewness. However, the slope of $P(x)$ at the suppressed end differs significantly from the former case.

Wrong normalization. Lastly, we consider the case in which $P(x)$ was calculated from a posterior with wrong normalization,

$$\begin{aligned} P^\epsilon(s|d) &= \frac{1}{\sqrt{2\pi}\sigma(1+\epsilon)}\exp\left(-\frac{s_d^2}{2\sigma^2}\right), \\ x^\epsilon &= \frac{1}{2(1+\epsilon)}\left[1+\mathrm{erf}\left(\frac{s_d}{\sqrt{2}\sigma}\right)\right], \end{aligned} \tag{14}$$

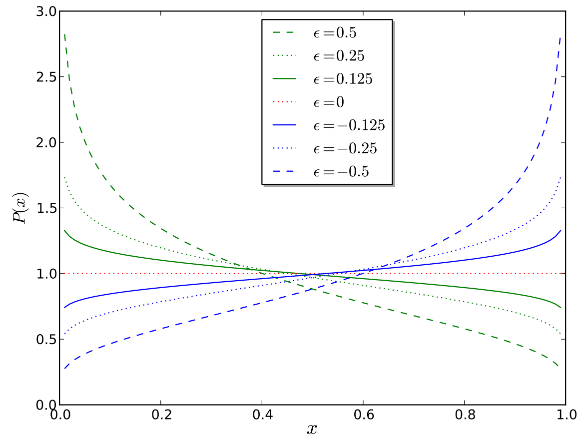

FIG. 2: Influence of an insufficient posterior on the DIP-distribution $P(x)$. The panel is showing the effect of calculating $P(x)$ from a posterior with wrong maximum position as described by Eq. (12).

which yields (Eq. (4))

$$P(x) = 1+\epsilon \qquad \text{for } x \in [0, 1-\epsilon]. \tag{15}$$

This means, the value of ϵ can be determined precisely from the x-interval.

A numerical example of an insufficient posterior. Next, we demonstrate the effects of insufficient posteriors with a numerical example. For this we generate mock data according to

$$d = s + n, \tag{16}$$

where s and n are zero-centered Gaussian random numbers with covariance $S = 1$ and $N = 0.1$, respectively. To reconstruct s optimally from the data we apply a *Wiener filter* [7] on d,

$$m = \underbrace{\left(S^{-1}+N^{-1}\right)^{-1}}_{=:D} N^{-1}d. \tag{17}$$

After that, the posterior for s is given by

$$P(s|d) = \mathcal{G}(s-m, D). \tag{18}$$

To investigate the accuracy of our implementation we go through the DIP validation procedure. For that purpose we sample s_{gen}-values from the distribution $\mathcal{G}(s, S)$. Next, we generate data according to Eq. (16) and calculate a posterior curve according to Eq. (18). Subsequently, we numerically determine the posterior probability for $s < s_{\mathrm{gen}}$, which is denoted by x. Now this procedure is repeated 500 times to sample $P(x)$.

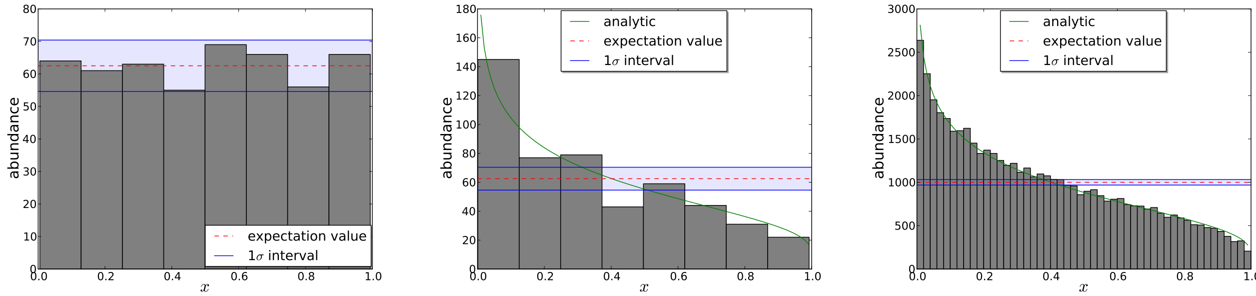

FIG. 3: Distributions of the numerically calculated x-values. The left (middle, right) histogram shows the unnormalized distribution of 500 (500, 50000) x-values within eight (eight, fifty) bins as calculated from the posterior with correct (wrong, wrong) maximum position. The standard deviation interval (1σ) around the expectation value as calculated from Poissonian statistics is also shown.

In order to demonstrate the effect of an insufficient posterior we falsely include a wrong maximum position with $\epsilon = 0.15$, i.e. our wrong test-posterior is given by

$$P^{\epsilon=0.15}(s|d) = \mathcal{G}(s - m - 0.15, D), \qquad (19)$$

and apply the validation procedure once again. Fig. 3 shows the distributions of x for the correct and incorrect posterior.
The results are in agreement with the analytical considerations. Another application of this validation approach in cosmology and its implications is given in [5]. There, a new way to calculate the posterior for the local primordial non-Gaussianity parameter $f_{\rm nl}$ from Cosmic Microwave Background observations is presented and validated.

With the help of the introduced tools of error-diagnosis it is possible to detect not only the presence of the mistake but also to get an indication of its nature.

Outlook. Although we have presented the DIP-test in one dimension ($s \in \mathbb{R}$) this approach can be extended to arbitrary dimensions ($t \in \mathbb{R}^m,\ m \in \mathbb{N}$) by mapping this multi-dimensional posterior $P(t|d)$ onto one dimension, $s = s(t) \in \mathbb{R}$. Now it is possible to apply the DIP-test for the remaining coordinate, $P(s|d)$. Because there are infinitely many ways to perform the mapping, $t \mapsto s = s(t)$, a suite of tests can be constructed to probe $P(t|d)$ in various ways. A combination of these tests then yields a multi-dimensional posterior test.

Furthermore, it is theoretically possible to do not only a qualitative error diagnosis, but also a quantitative study. One possibility is to consider the intersection of the distribution $P(x)$ of an insufficient posterior with the expectation value, $P(x) = 1$, which encodes (in combination with the shape and slope of $P(x)$) the value of ϵ. However, in reality there are combinations of different error types and numerically determined distributions are not as precise as the theoretical ones so that one might want to construct a Bayesian test for this.

We leave the development of fully automated error detection and classification methods for future work. Inspection of the results of the DIP-test by eye is already a powerful way to diagnose posterior imperfections, as we show in [5].

Acknowledgments. We want to thank Rishi Khatri for useful discussions.

[1] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, et al., ArXiv e-prints (2013), 1303.5076.

[2] S. R. Cook, A. Gelman, and D. B. Rubin, Journal of Computational and Graphical Statistics **15**, 675 (2006).

[3] J. Geweke, Journal of the American Statistical Association **99**, 799 (2004).

[4] T. Bayes, Phil. Trans. of the Roy. Soc. **53**, 370 (1763).

[5] S. Dorn, N. Opperman, R. Khatri, M. Selig, and T. A. Enßlin, ArXiv e-prints (2013).

[6] D. B. Owen, Ann. Math. Statist. **27**, 1075 (1956).

[7] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series* (New York: Wiley, 1949), ISBN 9780262730051.