On the relationship between a Gamma distributed precision parameter and the associated standard deviation in the context of Bayesian parameter inference

Manuel M. Eichenlaub*

School of Engineering, University of Warwick, United Kingdom

Abstract

In Bayesian inference, an unknown measurement uncertainty is often quantified in terms of a Gamma distributed precision parameter, which is impractical when prior information on the standard deviation of the measurement uncertainty shall be utilised during inference. This paper thus introduces a method for transforming between a gamma distributed precision parameter and the distribution of the associated standard deviation. The proposed method is based on numerical optimisation and shows adequate results for a wide range of scenarios.

Keywords: Gamma distribution, measurement uncertainty, Bayesian inference

1 Introduction

In the context of Bayesian parameter inference, it is common to model the error associated with the observed data as follows [1]:

\[ y(t) = g(\cdot) + \varepsilon \quad \text{with} \quad \varepsilon \sim \mathcal{N}(0, p^{-1}), \]

where \( y(t) \) is the observed data, \( g(\cdot) \) the observation function and \( \varepsilon \) the Gaussian distributed measurement error with zero mean and precision \( p \). This precision is commonly described as being Gamma distributed with shape and rate parameters \( a \) and \( b \), respectively

\[ p \sim \mathcal{G}(a, b) \quad \text{for} \quad p, a, b > 0 \]

The values of \( a \) and \( b \) define the probability density function (PDF) over \( p \) and are updated during parameter inference from a prior distribution of \( p \), defined by \( a_0 \) and \( b_0 \). The use of a Gamma distribution over \( p \) is justified by the fact that like the precision, the Gamma distribution is defined over positive values only. It furthermore forms a conjugate prior to the Gaussian distributed likelihood, therefore leading to analytically tractable posterior distributions and update rules [1][2]. An example of this approach can be found in a variational Bayesian method for the identification

*email: m.eichenlaub@warwick.ac.uk
stochastic nonlinear models [1, 3].

The prior for \( p \) is often chosen to be weak and uninformative [4]. However, in a number of practical applications, the collection of data is a known process and information on the measurement error \( \varepsilon \) can be found in the literature. Here, the measurement error is commonly quantified in the form of the standard deviation \( s \). An example for this would be the coefficient of variation of a certain immunoassay, e.g. of insulin. This information on \( s \) can therefore be used to specify the prior PDF over \( p \). Additionally, it is useful to allow the interpretation of the posterior distribution over \( p \) in terms of \( s \). This paper therefore introduces a method for the forwards and backwards transformation between the PDFs over \( p \) and \( s \).

2 Methodology

2.1 Transformation from \( p \) to \( s \)

The Gamma distribution over \( p \) is defined by the following PDF of shape and rate parameters \( a \) and \( b \), respectively [5],

\[
f_p(p|a,b) = \frac{b^a}{\Gamma(a)} p^{a-1} \exp(-pb) \quad \text{for} \quad p, a, b > 0,
\]

where \( \Gamma(\cdot) \) is the Gamma function. The mean and variance of this Gamma PDF are given by [5]

\[
\mathbb{E}[p]_{f_p} = \frac{a}{b} \quad \text{and} \quad \text{Var}[p]_{f_p} = \frac{a}{b^2}.
\]

The standard deviation \( s \) of the measurement error \( \varepsilon \) and its precision \( p \) are related as follows:

\[
s = \frac{1}{\sqrt{p}}
\]

In order to determine the PDF of \( s \) in terms of \( a \) and \( b \) the following theorem is used. If \( f_x \) is a PDF over the random variable \( x \) and the mapping \( y = h(x) \) is introduced, then the PDF over the random variable \( y \) is given by [5]:

\[
f_y(y) = f_x(h^{-1}(y)) \left| \frac{dh^{-1}(y)}{dy} \right|
\]

Defining \( s = h(p) = 1/\sqrt{p} \) from expression (5) and therefore \( h^{-1}(s) = 1/s^2 \), expression (6) can be used to determine the PDF \( f_s \) over \( s \) as follows:

\[
f_s(s|a, b) = f_p\left(\frac{1}{s^2}|a, b\right) \left| \frac{d}{ds} \frac{1}{s^2} \right|
\]

\[
= \frac{b^a}{\Gamma(a)} \left( \frac{1}{s^2} \right)^{a-1} \exp\left( -\frac{b}{s^2} \right) \frac{2}{s^3}
\]

\[
= \frac{2b^a}{\Gamma(a)} s^{-2a-1} \exp\left( -\frac{b}{s^2} \right)
\]
Using symbolic computation, this new probability distribution can be characterized by the following expression for the mean $\mu_s$ and the standard deviation $\sigma_s$, valid for $a > 1$:

$$\mu_s = \mathbb{E}[s]_{f_s} = \sqrt{b} \frac{\Gamma(a - \frac{1}{2})}{\Gamma(a)}.$$  

$$\sigma_s^2 = \text{Var}[s]_{f_s} = b \left[ \frac{1}{a-1} - \frac{\Gamma(a - \frac{1}{2})^2}{\Gamma(a)^2} \right].$$  

To facilitate the numerical calculation, the logarithm of the Gamma function $\log \Gamma(\cdot)$ is used instead of the fast growing Gamma function itself. This modifies expressions (8) to give

$$\mu_s = \sqrt{b} \exp \left[ \log \Gamma(a - \frac{1}{2}) - \log \Gamma(a) \right]$$

$$\sigma_s^2 = b \left[ \frac{1}{a-1} - \exp \left[ \log \Gamma(a - \frac{1}{2})^2 - \log \Gamma(a)^2 \right] \right].$$  

An example of the PDFs over $p$ and $s$ is provided in Figure 1, where it is demonstrated that the mean of $f_p$ does not simply transform into the mean of $f_s$ by applying the mapping $s = 1/\sqrt{p}$.

![Figure 1: Examples of the two PDFs of $f_p$ and $f_s$ for $a = b = 2$. The dashed vertical lines display the values of the respective means.](image_url)

### 2.2 Transformation from $s$ to $p$

Expressions (9) allows the interpretation of the posterior PDF of $p$, specified by $a$ and $b$ using the corresponding distribution over $s$. To specify the prior distribution over $p$ based on a chosen prior PDF over $s$, which in turn can be based on existing information, the following procedure is introduced. Defining $\mu_0$ and $\sigma_0$ characterising the the prior PDF over $s$, the goal is to calculate the associated values for $a_0$ and $b_0$, characterising the prior PDF over $p$. First, the following substitution is defined:

$$S(a) = \frac{\Gamma(a - \frac{1}{2})^2}{\Gamma(a)^2} = \exp \left[ \log \Gamma(a - \frac{1}{2})^2 - \log \Gamma(a)^2 \right].$$  

(10)
This is followed by the combination and reformulation of the expressions (9) into

\[ D(a) = \frac{\mu_0^2}{S(a)} - \frac{\sigma_0^2}{a - 1} - S(a) = 0. \]  

(11)

This eliminates \( b \) and makes it possible to find \( a_0 \) by solving the equation \( D(a_0) = 0 \) and subsequently calculating \( b_0 \) using:

\[ b_0 = \frac{\mu_0^2}{S(a_0)}. \]  

(12)

To find \( a_0 \), expression (11) is reformulated into a constrained numerical minimisation task:

\[ a_0 = \arg \min_a \left( \log \left[ D(a)^2 + 1 \right] \right) \quad \text{for} \quad a > 1. \]  

(13)

The square operation and addition of one within the logarithm ensures that the objective function is always positive except for \( \log[D(a_0)^2 + 1] \), where the expression is zero. The logarithm facilitates the numerical calculations as the values of only \( D(a)^2 \) would grow rapidly as \( a \) increases. An example of the objective function for different values of \( \mu_0 \) and \( \sigma_0 \) is given in Figure 2, demonstrating a clear minimum of the objective function at \( a_0 \).

Figure 2: Examples of the objective function \( \log[D(a)^2 + 1] \) for differing values of \( \mu_0 \) and \( \sigma_0 \). The minima correspond to the values of \( a_0 \), where \( D(a_0) = 0 \). The dashed lines give the respective values of the upper bound estimation of \( \hat{a}_0 \).

This minimisation can be numerically accomplished using an appropriate constrained implementation of an optimisation technique, e.g. \texttt{fminbnd} in MATLAB or \texttt{minimize_scalar} in Scipy. Since the expression (9) is only valid for \( a > 1 \) the lower bound is set to 1. To specify the upper bound based on the given values of \( \mu_0 \) and \( \sigma_0 \), the function \( S(a) \) from expression (10) is approximated with the first two terms of its series expansion for \( a \to \infty \), giving:

\[ \hat{S}(a) = \frac{1}{a} + \frac{3}{4a^2}. \]  

(14)
With this approximation, the minimum \( \hat{a}_0 \) of \( D(a) \) can be found analytically

\[
\hat{a}_0 = \frac{1}{8} \left[ 1 + \sqrt{49 + \frac{\mu_0^4}{\sigma_0^4} + 50 \frac{\mu_0^2}{\sigma_0^2} + \frac{\mu_0^2}{\sigma_0^2}} \right],
\]

which is used as the upper bound for the constrained minimisation (see Figure 2). Expressions (14) - (15) were found using symbolic calculation. This procedure now allows the calculation of \( a_0 \) and subsequently \( b_0 \) with expression (12) based on \( \mu_0 \) and \( \sigma_0 \), thereby specifying the prior distribution over \( p \). All symbolic calculations were done using Mathematica and the respective code is available online (https://github.com/manueich/Noise_Gamma).

2.3 Validation

In order to assess the accuracy of the numerical calculations when transforming from \( s \) to \( p \), the method was implemented in MATLAB 2020a with the function \textit{fminbnd} and Python with the function \textit{minimize_scalar} using identical optimisation settings. The respective code is available online (https://github.com/manueich/Noise_Gamma).

As a test, a wide range of possible values for \( \mu \) and \( \sigma \) are transformed into values for \( a \) and \( b \) and subsequently back-transformed into \( \mu \) and \( \sigma \) using expressions (9). These results can then be compared to the starting values of \( \mu \) and \( \sigma \). A test is considered as passed if the original values for \( \mu \) and \( \sigma \) can be recovered with a relative error smaller than 1 \% on both parameters. For \( \mu \), a total of 1000 values logarithmically scaled on a range between \( 10^{-4} \) and \( 10^{4} \) are chosen. Values outside this range should not be encountered in practice. Subsequently, for each value of \( \mu \), a total of 1000 values for \( \sigma \) are logarithmically scaled on a range between \( 10^{-4} \cdot \mu \) and \( 10^{2} \cdot \mu \) are tested.

3 Results and discussion

The results of the validation procedure for both MATLAB and Python are displayed in Figure 3 and are very similar. Based on this the following cut-off values are proposed, which should ensure a robust calculation of \( a \) and \( b \) based on \( \mu \) and \( \sigma \).

- \( 2 \cdot 10^{-3} < \mu < 10^{4} \)
- \( 3 \cdot 10^{-3} < \frac{\sigma}{\mu} < 50 \)

These ranges should cover a large number of possible values that could be encountered in practice.
Figure 3: Results of the validation procedure for (a) MATLAB and (b) Python. Red indicates a failed and green indicates a passed test. The solid black lines give the suggested cut-off values.

References

[1] J. Daunizeau, K. J. Friston, and S. J. Kiebel. “Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models”. In: Physica D: Nonlinear Phenomena 238.21 (2009), pp. 2089–2118.

[2] Christopher M. Bishop. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag, 2006.

[3] Jean Daumizeau, Vincent Adam, and Lionel Rigoux. “VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data”. In: PLoS Comput Biol 10.1 (2014).

[4] Andrew Gelman. “Prior distributions for variance parameters in hierarchical models”. In: Bayesian Analysis 1 (2006), pp. 1–19.

[5] I. Florescu and C.A. Tudor. Handbook of Probability. Wiley, 2014.