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

Manuel M. Eichenlaub[1,2*]

[1]School of Engineering, University of Warwick, United Kingdom
[2]Human Metabolism Research Unit, University Hospitals Coventry and Warwickshire, United Kingdom

09/05/2019

In the context of Bayesian parameter inference, it is common practice to model the error associated with the observed data as additive Gaussian distributed white noise with zero mean and certain precision (inverse variance) [1]:

$$y(t) = g(\cdot) + \epsilon \quad \text{with} \quad \epsilon \sim \mathcal{N}(0, p^{-1}), \tag{1}$$

where $y(t)$ is the observed data, $g(\cdot)$ the observation function, ϵ the Gaussian distributed measurement error and p its precision. In a fully Bayesian treatment of all parameters, this precision is often modelled as being Gamma distributed with shape and rate parameters a and b, respectively.

$$p \sim Ga(a, b) \quad \text{for} \quad a, b > 0 \tag{2}$$

a and b are so called hyperparameters of p and are used to define its prior distribution. They are updated during parameter inference and subsequently also define the posterior distribution of p. 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 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 ϵ can be found in literature. Here, it is common practice to quantify the measurement error in the form of e.g. a mean absolute deviation, a mean absolute relative deviation or a coefficient of variation. All of these quantities typically provide information on the standard deviation of ϵ, not its precision p. This leads to two important questions:

*email: m.eichenlaub@warwick.ac.uk

- How can the posterior distribution of p be interpreted in the context of the information on the measurement error found in literature?
- How can we characterize the prior distribution over p based on the information on the measurement error?

In this work we answer those questions by presenting a method for relating a Gamma distributed precision parameter to the associated standard deviation by utilizing the inferred hyperparameters a and b.

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

$$f_p(p|a,b) = \frac{b^a}{\Gamma(a)} p^{a-1} \exp(-pb) \quad \text{for} \quad p > 0, \tag{3}$$

where $\Gamma(\cdot)$ is the Gamma function. The standard deviation s of the measurement error ϵ and its precision p are related as follows:

$$s = \frac{1}{\sqrt{p}} \tag{4}$$

In order to determine the distribution of s from the distribution of p using relationship (4), we can use the following theorem [5].
If f_x is a PDF over the random variable x and we assume the mapping of $y = h(x)$, the PDF over the random variable y is given by:

$$f_y(y) = f_x(h^{-1}(y)) \left| \frac{\mathrm{d}h^{-1}(y)}{\mathrm{d}y} \right| \tag{5}$$

If we define $s = h(p) = 1/\sqrt{p}$ from (4) and therefore $h^{-1}(s) = 1/s^2$, we can use (5) to determine the PDF f_s over s as follows:

$$\begin{aligned} f_s(s|a,b) &= f_p(\frac{1}{s^2}|a,b) \left| \frac{\mathrm{d}}{\mathrm{d}s} \frac{1}{s^2} \right| \\ &= \frac{b^a}{\Gamma(a)} \left(\frac{1}{s^2} \right)^{a-1} \exp\left(-\frac{b}{s^2} \right) \frac{2}{s^3} \\ &= \frac{2b^a}{\Gamma(a)} s^{-2a-1} \exp\left(-\frac{b}{s^2} \right) \end{aligned} \tag{6}$$

This new probability distribution can be characterized by the following expression for the mean μ_s and the standard deviation σ_s, valid for $a > 1$:

$$\mu_s = \mathbb{E}[s]_{f_s} = \sqrt{b} \frac{\Gamma(a - \frac{1}{2})}{\Gamma(a)}, \tag{7}$$

$$\sigma_s^2 = \mathbb{E}[s^2]_{f_s} - \mathbb{E}[s]_{f_s}^2 = b \left[\frac{1}{a-1} - \frac{\Gamma(a - \frac{1}{2})^2}{\Gamma(a)^2} \right], \tag{8}$$

where $\mathbb{E}[s]_{f_s}$ is the expected value of s with respect to f_s. The mode of this new PDF over s can be specified as well for $a > 1$:

$$m_s = \sqrt{\frac{2b}{1+2a}} \tag{9}$$

An example of the PDFs for f_p and f_s is displayed in Figure 1.

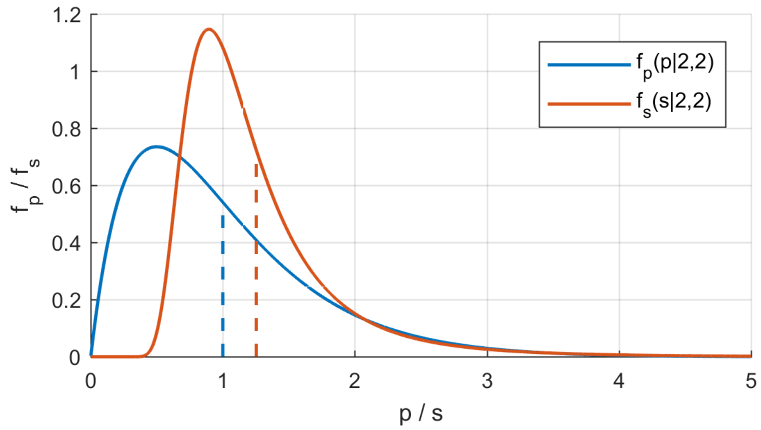

Figure 1: Examples of the two PDFs of f_p and f_s for $a = b = 2$. The dashed vertical lines display the values of the respective means.

In order to facilitate the numerical calculation of these values, it is helpful to use the natural logarithm of the Gamma function $\ln\Gamma(\cdot)$ instead of the fast growing Gamma function itself. This alters expressions (7) and (8) as follows:

$$\mu_s = \sqrt{b}\exp\left[\ln\Gamma(a-\frac{1}{2}) - \ln\Gamma(a)\right] \tag{10}$$

$$\sigma_s^2 = b\left[\frac{1}{a-1} - \exp\left[\ln\Gamma(a-\frac{1}{2})^2 - \ln\Gamma(a)^2\right]\right] \tag{11}$$

These expressions are the answer to the first question and provide a straightforward way of calculating the sufficient statistics of the inferred measurement error standard deviation s from the hyperparameters a and b.

This leaves us to find the answer to the second question, which poses the inverse problem of finding the values for a_0 and b_0, given μ_s and σ_s for the definition of a prior distribution for the precision p.

For that system of equations defined by (10) and (11) have to solved for a and b, respectively. We start by introducing the substitution:

$$S(a) = \frac{\Gamma(a-\frac{1}{2})^2}{\Gamma(a)^2} = \exp\left[\ln\Gamma(a-\frac{1}{2})^2 - \ln\Gamma(a)^2\right], \tag{12}$$

followed by the combination and reformulation of (10) and (11) into the following expression:

$$D(a) = \frac{\mu_s^2}{S(a)} - \frac{\sigma_s^2}{\frac{1}{a-1} - S(a)} \tag{13}$$

It is now possible to find a_0 by solving the equation $D(a_0) = 0$ and subsequently finding b_0 using the following expression:

$$b_0 = \frac{\mu_s^2}{S(a_0)} \tag{14}$$

To the author's knowledge, there is no exact analytical solution for a_0. However we can find a numerical approximation by reformulation this problem into a constrained optimization task:

$$a_0 = \min_a \ln\left[D(a)^2 + 1\right] \quad \text{for} \quad a > 1 \tag{15}$$

The natural logarithm facilitates numerical calculations as the values of $D(a)^2$ grow rapidly as a increases. The square operation and subtraction of one within the logarithm ensures a minimum at $D(a_0) = 0$ and also $\ln\left[D(a_0)^2 + 1\right] = 0$. Examples of the function $\ln\left[D(a)^2 + 1\right]$ for different values of μ_s and σ_s are displayed in Figure 2.

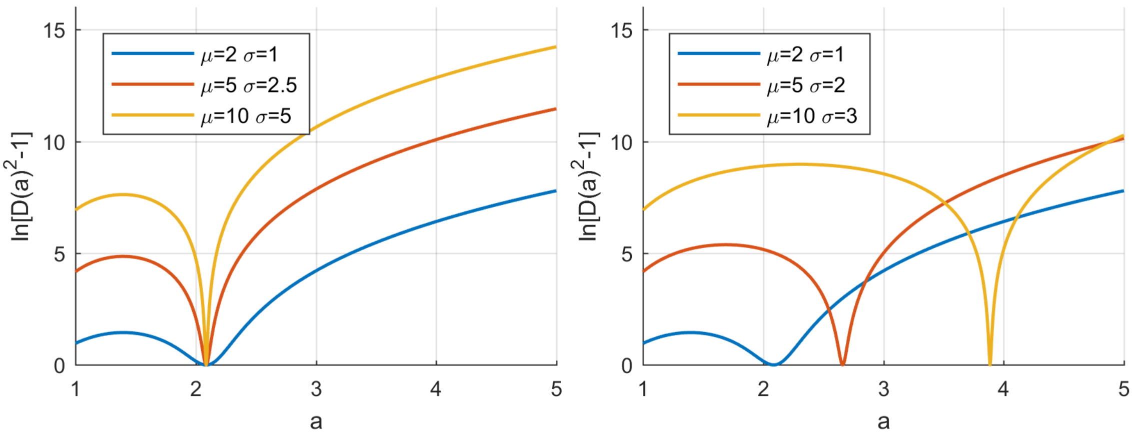

Figure 2: Examples of the function $\ln\left[D(a)^2 + 1\right]$ for different values of μ_s and σ_s. In the left figure, the ratios between μ_s and σ_s are kept constant, leading to the same minimum.

In order to solve this constraint minimization problem, an algorithm based on a interior-point approach was used. In our case the was constraint to a search space $a > 1$. In order to set an initial value based on the given parameters μ_s and σ_s the function $S(a)$ in (12) can be approximated with the first two terms of its power series expansion for $a \to \infty$:

$$\hat{S}(a) = \frac{1}{a} + \frac{3}{4a^2} + \mathcal{O}\left(\frac{1}{a}\right)^{5/2} \tag{16}$$

Using this approximation we can find an exact solution for $D(\hat{a}_0) = 0$:

$$\hat{a}_0 = \frac{1}{8}\left[1 + \sqrt{49 + \frac{\mu_s^4}{\sigma_s^4} + 50\frac{\mu_s^2}{\sigma_s^2}} + \frac{\mu_s^2}{\sigma_s^2}\right] \tag{17}$$

This expression for $\hat{a}_0$ is subsequently used as a initial value.

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 Daunizeau, 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.