Journal of Behavioral Data Science, 2021, *1* (2), 1–8.
DOI:https://doi.org/10.35566/jbds/v1n2/p2

A Note on Wishart and Inverse Wishart Priors for Covariance Matrix

Zhiyong Zhang[0000−0003−0590−2196]

University of Notre Dame

Abstract. For inference involving a covariance matrix, inverse Wishart priors are often used in Bayesian analysis. To help researchers better understand the influence of inverse Wishart priors, we provide a concrete example based on the analysis of a two by two covariance matrix. Recommendations are provided on how to specify an inverse Wishart prior.

Keywords: Wishart distribution · inverse Wishart distribution · prior distribution · covariance matrix

In Bayesian analysis, an inverse Wishart (IW) distribution is often used as a prior for the variance-covariance parameter matrix (e.g., Barnard, McCulloch,x& Meng, 2000; Gelmanxet al., 2014; Leonard, Hsu,xet al., 1992). The IW prior is very popular because it is conjugate to normal data. For best illustration, consider a multivariate normal (MN) variable. Let $\mathbf{X} = (X_1, X_2, \ldots, X_p)$ denote a vector of p variables

$$\mathbf{X}|\mathbf{\Sigma} \sim MN(\mathbf{0}, \mathbf{\Sigma})$$

with the mean vector $\boldsymbol{\mu} = \mathbf{0}$ and the variance-covariance matrix $\mathbf{\Sigma}$. The density function is

$$p(\mathbf{x}|\mathbf{\Sigma}) = (2\pi)^{-p/2}|\mathbf{\Sigma}|^{-1/2}\exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}\right).$$

Given a sample $\mathbf{D} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ with n being the sample size, the likelihood function for $\mathbf{\Sigma}$ is

$$\begin{aligned} L(\mathbf{\Sigma}|\mathbf{D}) \propto p(\mathbf{D}|\mathbf{\Sigma}) &\propto |\mathbf{\Sigma}|^{-n/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\mathbf{x}_i^T\mathbf{\Sigma}^{-1}\mathbf{x}_i\right) \\ &= |\mathbf{\Sigma}|^{-n/2}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^T\mathbf{\Sigma}^{-1}\right)\right] \\ &= |\mathbf{\Sigma}|^{-n/2}\exp\left[-\frac{n}{2}\mathrm{tr}(\mathbf{S}\mathbf{\Sigma}^{-1})\right], \end{aligned}$$

where $\mathbf{S} = \sum_i^n \mathbf{x}_i\mathbf{x}_i^T/n$ is the biased sample covariance matrix (the sample is centered at 0). Note that this is also the maximum likelihood estimate of $\mathbf{\Sigma}$. To

get the posterior distribution of $\mathbf{\Sigma}$ for Bayesian inference, one needs to specify a prior distribution $p(\mathbf{\Sigma})$ for it. With the prior, the posterior distribution can be obtained through the Bayes' Theorem:

$$p(\mathbf{\Sigma}|\mathbf{D}) = \frac{p(\mathbf{D}|\mathbf{\Sigma})p(\mathbf{\Sigma})}{p(\mathbf{D})}.$$

1 The Inverse Wishart Prior

The most commonly used prior for $\mathbf{\Sigma}$ is probably the inverse Wishart conjugate prior. The density function of an inverse Wishart distribution $IW(\mathbf{V}, m)$ with the scale matrix $\mathbf{V}$ and the degrees of freedom m for a $p \times p$ variance-covariance matrix $\mathbf{\Sigma}$ is

$$p(\mathbf{\Sigma}) = \frac{|\mathbf{V}|^{m/2}|\mathbf{\Sigma}|^{-(m+p+1)/2}\exp\left[-\mathrm{tr}(\mathbf{V}\mathbf{\Sigma}^{-1})/2\right]}{2^{mp/2}\Gamma(m/2)}.$$

The inverse Wishart distribution is a multivariate generalization of the inverse Gamma distribution. The mean of it is

$$E(\mathbf{\Sigma}) = \frac{\mathbf{V}}{m-p-1} \tag{1}$$

and the variance of each element of $\mathbf{\Sigma} = (\sigma_{ij})$ is

$$Var(\sigma_{ij}) = \frac{(m-p+1)v_{ij}^2 + (m-p-1)v_{ii}v_{jj}}{(m-p)(m-p-1)^2(m-p-3)}.$$

Especially,

$$Var(\sigma_{ii}) = \frac{2v_{ii}^2}{(m-p-1)^2(m-p-3)}. \tag{2}$$

With an inverse Wishart prior $IW(\mathbf{V}_0, m_0)$ based on known $\mathbf{V}_0$ and m_0, the posterior distribution of $\mathbf{\Sigma}$ is

$$\begin{aligned} p(\mathbf{\Sigma}|\mathbf{D}) &\propto p(\mathbf{D}|\mathbf{\Sigma})p(\mathbf{\Sigma}) \\ &= |\mathbf{\Sigma}|^{-n/2}\exp\left[-\frac{n}{2}\mathrm{tr}(\mathbf{S}\mathbf{\Sigma}^{-1})\right]|\mathbf{\Sigma}|^{-(m_0+p+1)/2}\exp\left[-\mathrm{tr}(\mathbf{V}_0\mathbf{\Sigma}^{-1})/2\right] \\ &= |\mathbf{\Sigma}|^{-(n+m_0+p+1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[(n\mathbf{S}+\mathbf{V}_0)\,\mathbf{\Sigma}^{-1}\right]\right\}. \end{aligned}$$

From it, we can get the posterior distribution for $\mathbf{\Sigma}$, also an inverse Wishart distribution:

$$\mathbf{\Sigma}|\mathbf{D} \sim IW\left(n\mathbf{S}+\mathbf{V}_0, n+m_0\right) = IW(\mathbf{V}_1, m_1) \tag{3}$$

with the updated scale matrix and degrees of freedom.

1.1 Information in an inverse Wishart prior

The posterior mean of $\mathbf{\Sigma}$ is

$$\begin{aligned} E(\mathbf{\Sigma}|\mathbf{D}) &= \frac{n\mathbf{S} + \mathbf{V}_0}{n + m_0 - p - 1} \\ &= \frac{n}{n + m_0 - p - 1}\mathbf{S} + \left(1 - \frac{n}{n + m_0 - p - 1}\right)\frac{\mathbf{V}_0}{m_0 - p - 1}. \end{aligned} \quad (4)$$

Therefore, the posterior mean is a weighted average of the sample covariance matrix $\mathbf{S}$ and the prior mean $\mathbf{V}_0/(m_0 - p - 1)$. When the sample size $n \to \infty$, the posterior mean approaches the sample mean given fixed m_0 and p.

The information in a prior can be connected to data. For example, if we specify the prior $IW(\mathbf{V}_0, m_0)$ as $\mathbf{V}_0 = n_0\mathbf{S}$ and $m_0 = n_0$, then the informative in the prior is equivalent to n_0 participants in the sample. Note that if we set $\mathbf{V}_0 = (m_0 - p - 1)\mathbf{S}$, then $E(\mathbf{\Sigma}|\mathbf{D}) = \mathbf{S}$, meaning the posterior mean is the same as the sample covariance matrix.

2 Precision Matrix and the Wishart Prior

In practice, the BUGS program is probably the most widely used software for Bayesian analysis (e.g., Lunn, Jackson, Best, Thomas,x& Spiegelhalter, 2012; Ntzoufras, 2009). BUGS uses the precision matrix, defined as the inverse of the covariance matrix, to specify the multivariate normal distribution. Let $\mathbf{P} = \mathbf{\Sigma}^{-1}$, then the normal density function can be written as

$$p(\mathbf{x}|\mathbf{P}) = (2\pi)^{-p/2}|\mathbf{P}|^{1/2}\exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x}\right).$$

The use of the precision matrix has the computational advantage by avoiding the inverse of matrix in the density calculation in certain situations.

For the precision matrix $\mathbf{P}$, a Wishart prior $W(\mathbf{U}_0, w_0)$ with the scale matrix $\mathbf{U}_0$ and degrees of freedom w_0 is used (e.g., Lunnxet al., 2012). The density function of the prior is

$$p(\mathbf{P}) = \frac{|\mathbf{P}|^{(w_0 - p - 1)/2}\exp\left[-\mathrm{tr}(\mathbf{U}_0^{-1}\mathbf{P})/2\right]}{2^{w_0 p/2}\Gamma(w_0/2)|\mathbf{U}_0|^{w_0/2}}.$$

Given the sample $\mathbf{D} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, the posterior distribution of $\mathbf{P}$ is

$$\begin{aligned} p(\mathbf{P}|\mathbf{D}) &\propto \prod_{i=1}^{n}\left[|\mathbf{P}|^{1/2}\exp\left(-\frac{1}{2}\mathbf{x}_i^T\mathbf{P}\mathbf{x}_i\right)\right]|\mathbf{P}|^{(w_0 - p - 1)/2}\exp\left[-\mathrm{tr}(\mathbf{U}_0^{-1}\mathbf{P})/2\right] \\ &= |\mathbf{P}|^{(n + w_0 - p - 1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[(n\mathbf{S} + \mathbf{U}_0^{-1})\mathbf{P}\right]\right\}. \end{aligned}$$

Therefore, the posterior is also a Wishart distribution $W(\mathbf{U}_1, w_1)$ with $\mathbf{U}_1 = \left(n\mathbf{S} + \mathbf{U}_0^{-1}\right)^{-1}$ and $w_1 = n + w_0$. The posterior mean of $\mathbf{P}$ is

$$E(\mathbf{P}|\mathbf{D}) = w_1\mathbf{U}_1 = (n + w_0)\left(n\mathbf{S} + \mathbf{U}_0^{-1}\right)^{-1}.$$

Based on the relationship between Wishart and inverse Wishart distributions (Mardia, Bibby,x& Kent, 1982),

$$\mathbf{\Sigma}|\mathbf{D} = \mathbf{P}^{-1}|\mathbf{D} \sim IW(\mathbf{U}_1^{-1}, w_1) = IW(n\mathbf{S} + \mathbf{U}_0^{-1}, n + w_0). \tag{5}$$

The posterior mean of $\mathbf{\Sigma}$ is

$$E(\mathbf{\Sigma}|\mathbf{D}) = \frac{\mathbf{U}_1^{-1}}{w_1 - p - 1} = \frac{n\mathbf{S} + \mathbf{U}_0^{-1}}{n + w_0 - p - 1}. \tag{6}$$

Comparing the posterior distributions in Equation (3) and (5), giving an inverse Wishart distribution $IW(\mathbf{V}_0, m_0)$ prior to the covariance matrix $\mathbf{\Sigma}$ is the same as giving a Wishart distribution $W(\mathbf{V}_0^{-1}, m_0)$ prior to the precision matrix $\mathbf{P} = \mathbf{\Sigma}^{-1}$. However, note that

$$[E(\mathbf{P}|\mathbf{D})]^{-1} = \frac{n\mathbf{S} + \mathbf{U}_0^{-1}}{n + w_0} \neq E(\mathbf{\Sigma}|\mathbf{D}) = \frac{n\mathbf{S} + \mathbf{U}_0^{-1}}{n + w_0 - p - 1}.$$

Therefore, one cannot simply invert the posterior mean of the precision matrix to get the posterior mean of the covariance matrix.

3 Numerical Examples

For illustration, we look at a concrete experiment. Suppose we have a sample of size $n = 100$ with the sample covariance matrix ($p = 2$)

$$\mathbf{S} = \begin{pmatrix} 5 & 2 \\ 2 & 10 \end{pmatrix}.$$

The aim is to estimate $\mathbf{\Sigma}$ through Bayesian method. We now consider the use of different priors and evaluate their influence. Given the connection between the Wishart and inverse Wishart distributions, we focus our discussion on the specification of an inverse Wishart prior for the covariance matrix $\mathbf{\Sigma}$.

3.1 Priors based on an identity scale matrix

For an inverse Wishart prior $IW(\mathbf{V}_0, m_0)$, we need to specify its scale matrix and degrees of freedom. In practice, an identity matrix has been frequently used as the scale matrix. Therefore, we first set $\mathbf{V}_0 = \mathbf{I}$ and vary the degrees of freedom by letting $m_0 = 2, 5, 10, 50, 100$. Note that when $m_0 = 2$, the prior is not a proper distribution but the posterior is still a proper distribution. The mean and variance of the posterior distribution are given in Table 1. First, when

$m_0 = 2$ or 5, the posterior means are close to the sample covariance matrix. With the increase of m_0, the posterior means become smaller and the posterior variances also become smaller. This can be easily explained by Equation (4) – the posterior mean is a weighted average between the sample mean and the prior mean. Take the element Σ_{11} as an example. From the data, $S_{11} = 5$. The mean of the inverse Wishart prior is $V_{0,11}/(m_0 - 3) = 1/(m_0 - 3)$. When $m_0 = 5$, the prior mean is 0.5 and when $m_0 = 100$, the prior mean is about 0.01. Furthermore, when $m_0 = 5$, the weight for the prior mean is about 0.05 but when $m_0 = 100$, the weight increases to about 0.5. Therefore, with the increase of m_0, the posterior mean is pulled towards the prior mean since the prior mean has a greater weight.

Table 1. Posterior inference of the covariance matrix parameter based on the inverse Wishart prior with the scale matrix specified based on an identity matrix.

		Mean					Variance				
	S	2	5	10	50	100	2	5	10	50	100
$IW(\mathbf{I}, m_0)$											
Σ_{11}	5	5.06	4.91	4.68	3.41	2.54	0.528	0.483	0.418	0.160	0.066
Σ_{12}	2	1.96	1.96	1.87	1.36	1.02	0.516	0.516	0.447	0.172	0.071
Σ_{22}	10	10.11	9.81	9.36	6.81	5.08	2.108	1.926	1.667	0.640	0.265
$IW[(m_0 - p - 1)\mathbf{I}, m_0]$											
Σ_{11}	5	5.04	4.92	4.74	3.72	3.03	0.524	0.484	0.428	0.191	0.094
Σ_{12}	2	1.96	1.96	1.87	1.36	1.02	0.518	0.518	0.454	0.194	0.091
Σ_{22}	10	10.09	9.82	9.41	7.12	5.57	2.100	1.930	1.687	0.700	0.318

In the above specification, since $\mathbf{V}_0 \equiv \mathbf{I}$, the prior mean also changes along the change of m_0. In practice, e.g., in sensitivity analysis, it can be helpful to fix the prior mean. To achieve this, one can set $\mathbf{V}_0 = (m_0 - p - 1)\mathbf{I}$. Therefore, when $m_0 = 5$, the scale matrix will be $2\mathbf{I}$, and when $m_0 = 100$, the scale matrix will be $m_0 = 97\mathbf{I}$. With such specification, the prior mean is always $\mathbf{I}$.

3.2 Priors with the scale matrix formed from data

Another way to specify the prior is to construct the scale matrix for the inverse Wishart distribution based on the sample data. Intuitively, we can set $\mathbf{V}_0 = \mathbf{S}$ and change m_0. From the top of Table 2, with the increase of m_0, the posterior mean deviates from the sample covariance matrix. This is again because that the prior mean becomes smaller with the increase of m_0 since the prior mean is equal to $\mathbf{S}/m_0$. To maintain the same prior mean while changing the information in the prior, we set $\mathbf{V}_0 = (m_0 - p - 1)\mathbf{S}$. With such specification, the prior mean is always $\mathbf{S}$ and the posterior mean is also $\mathbf{S}$ as we can see from the bottom part of Table 2. With the increase of the degrees of freedom, more information is supplied through the prior and we can observe the decrease in the posterior variance.

Table 2. Posterior inference of the covariance matrix parameter based on the priors with the scale matrix constructed from data.

		Mean					Variance				
	S	2	5	10	50	100	2	5	10	50	100
$IW(\mathbf{S}, m_0)$											
Σ_{11}	5	5.10	4.95	4.72	3.44	2.56	0.537	0.490	0.424	0.163	0.067
Σ_{12}	2	1.98	1.98	1.89	1.37	1.03	0.525	0.525	0.455	0.175	0.072
Σ_{22}	10	10.20	9.90	9.44	6.87	5.13	2.146	1.961	1.697	0.651	0.270
$IW[(m_0 - p - 1)\mathbf{S}, m_0]$											
Σ_{11}	5	5.00	5.00	5.00	5.00	5.00	0.515	0.500	0.476	0.345	0.256
Σ_{12}	2	2.00	2.00	2.00	2.00	2.00	0.536	0.536	0.510	0.370	0.276
Σ_{22}	10	10.00	10.00	10.00	10.00	10.00	2.062	2.000	1.905	1.379	1.026

3.3 Other types of specifications

We now consider several other types of specifications of the scale matrix to illustrate the influence of the prior. In all the the specifications, we maintain the same prior mean by setting the prior in the form of $IW[(m_0-p-1)\mathbf{V}_0, m_0]$. The priors considered and the associated posterior mean and variance are summarized in Table 3.

For prior P1, it assumes that Σ_{11} is 10 times of Σ_{22}, which is not consistent with the sample data. As expected, the posterior mean is pulled towards prior mean with the increase of m_0. Notably, the variance of Σ_{11} does not monotonously decrease with the increase of m_0 as one might incorrectly assume that the use of prior information will lead to more precise results. This is because the variance of the inverse Wishart distribution is related to its mean as shown in Equation (2), and the prior is not consistent with data.

For Priors P2, P3, P4, and the one at the bottom of Figure 2, the scale matrices have the same diagonal values and different off-diagonal values. Note that changing the values on the off-diagonals influences neither the posterior means nor variances on the diagonals, which can also be seen in Equations (1) and (2). As expected, changing the off-diagonal values influences both the posterior means and variances. However, the posterior variances are relatively stable.

3.4 Using priors for a precision matrix P

The influence of the priors on the precision matrix is the same as for the covariance matrix because of the connection of Wishart and inverse Wishart distribution – if $\mathbf{\Sigma} \sim IW(\mathbf{V}_0, m_0)$, $\mathbf{P} = \mathbf{\Sigma}^{-1} \sim W(\mathbf{V}_0^{-1}, m_0)$. If the prior $IW(\mathbf{I}, m_0)$ is specified for the covariance matrix, it is equivalent to use $W(\mathbf{I}, m_0)$ for the precision matrix. As discussed earlier, to maintain the same prior mean, we can use $IW[(m_0 - p - 1)\mathbf{I}, m_0]$ for $\mathbf{\Sigma}$. In this case, the prior for the precision matrix should be $W[\mathbf{I}/(m_0 - p - 1), m_0]$. Similarly, if we specify a prior for $\mathbf{\Sigma}$ based on the data using $IW[(m_0 - p - 1)\mathbf{S}, m_0]$, then the prior for the precision matrix would be $W[\mathbf{S}^{-1}/(m_0 - p - 1), m_0]$.

Table 3. Posterior inference of the covariance matrix parameter with additional specifications of inverse Wishart priors $IW[(m_0 - p - 1)\mathbf{V}_0, m_0]$.

		Mean					Variance				
	S	2	5	10	50	100	2	5	10	50	100
P1: $\mathbf{V}_0 = \begin{pmatrix} 10 & 0 \\ 0 & 1 \end{pmatrix}$											
Σ_{11}	5	4.95	5.10	5.33	6.60	7.46	0.505	0.520	0.541	0.601	0.571
Σ_{12}	2	1.96	1.96	1.87	1.36	1.02	0.535	0.535	0.507	0.335	0.217
Σ_{22}	10	10.09	9.82	9.41	7.12	5.57	2.100	1.930	1.687	0.700	0.318
P2: $\mathbf{V}_0 = \begin{pmatrix} 5 & -2 \\ -2 & 10 \end{pmatrix}$											
Σ_{11}	5	5.00	5.00	5.00	5.00	5.00	0.515	0.500	0.476	0.345	0.256
Σ_{12}	2	1.92	1.92	1.74	0.72	0.03	0.532	0.532	0.501	0.346	0.255
Σ_{22}	10	10.00	10.00	10.00	10.00	10.00	2.062	2.000	1.905	1.379	1.026
P3: $\mathbf{V}_0 = \begin{pmatrix} 5 & 0 \\ 0 & 10 \end{pmatrix}$											
Σ_{11}	5	5.00	5.00	5.00	5.00	5.00	0.515	0.500	0.476	0.345	0.256
Σ_{12}	2	1.96	1.96	1.87	1.36	1.02	0.534	0.534	0.505	0.355	0.260
Σ_{22}	10	10.00	10.00	10.00	10.00	10.00	2.062	2.000	1.905	1.379	1.026
P4: $\mathbf{V}_0 = \begin{pmatrix} 5 & -5 \\ -5 & 10 \end{pmatrix}$											
Σ_{11}	5	5.00	5.00	5.00	5.00	5.00	0.515	0.500	0.476	0.345	0.256
Σ_{12}	2	1.86	1.86	1.54	-0.24	-1.45	0.530	0.530	0.495	0.343	0.266
Σ_{22}	10	10.00	10.00	10.00	10.00	10.00	2.062	2.000	1.905	1.379	1.026

4 Discussion

Although not without issues, Wishart and inverse Wishart distributions are still commonly used prior distributions for Bayesian analysis involving a covariance matrix (Alvarez, Niemi,x& Simpson, 2014; Liu, Zhang,x& Grimm, 2016). As we have shown, the use of the inverse Wishart prior has the advantage of conjugate, which simplifies the posterior distribution. By using an inverse Wishart prior, the posterior distribution is also an inverse Wishart distribution given normally distributed data. The posterior mean can be conveniently expressed as a weighted average of the prior mean and the sample covariance matrix. The influence of the prior can also be clearly quantified.

When reliable information is available, an informative inverse Wishart prior can be constructed. For example, previous estimates on the covariance matrix could be available. In this situation, such covariance matrix estimates can be used to construct the scale matrix. If the variance estimates of the covariance matrix is also available, one can determine the degrees of freedom for the inverse Wishart prior based on the variance expression in Equation (2), which can be done using the R package discussed in the Appendix. The degrees of freedom based on each individual element may vary. The overall degrees of freedom for the inverse Wishart distribution can be determined based on the practical research question.

When no reliable information is available, an identity matrix has often been suggested to use as the scale matrix for the inverse Wishart distribution for the covariance matrix and Wishart distribution for the precision matrix (e.g., Congdon, 2014). But as one can see from the numerical example, how much information such a prior has is related to the covariance matrix. We believe a better way to specify an uninformative prior is to determine the scale matrix based on the sample covariance matrix. Therefore, we recommend the prior $IW[(m_0 - p - 1)\mathbf{S}, m_0]$. As for the precision matrix, one can use $W[\mathbf{S}^{-1}/(m_0 - p - 1), m_0]$.

Appendix

The R package wishartprior is developed and made available on GitHub to help understand the Wishart and inverse Wishart priors. The URL to the package is `https://github.com/johnnyzhz/wishartprior`. The package can be used to generate random numbers from an inverse Wishart distribution. It can calculate the mean and variance of Wishart and inverse Wishart distributions. Using the package, one can investigate the influence of priors.

References

Alvarez, I., Niemi, J., & Simpson, M. (2014). Bayesian inference for a covariance matrix. In *Anual conference on applied statistics in agriculture* (pp. 71–82). Retrieved from `arXiv:1408.4050`

Barnard, J., McCulloch, R., & Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, *10*, 1281–1311.

Congdon, P. (2014). *Applied bayesian modeling* (2nd ed.). John Wiley & Sons.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (2nd ed.). CRC press.

Leonard, T., Hsu, J. S., et al. (1992). Bayesian inference for a covariance matrix. *The Annals of Statistics*, *20*(4), 1669–1696. doi: https://doi.org/10.1214/aos/1176348885

Liu, H., Zhang, Z., & Grimm, K. J. (2016). Comparison of inverse wishart and separation-strategy priors for bayesian estimation of covariance parameter matrix in growth curve analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(3), 354–367. doi: https://doi.org/10.1080/10705511.2015.1057285

Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The bugs book: A practical introduction to bayesian analysis*. CRC Press.

Mardia, K., Bibby, J., & Kent, J. (1982). *Multivariate analysis*. Academic Press.

Ntzoufras, I. (2009). *Bayesian modeling using WinBUGS*. John Wiley & Sons.