A Characterization of the Bivariate Normal-Wishart Distribution

Dan Geiger
dang@cs.technion.ac.il

David Heckerman
heckerma@microsoft.com

November 1995

Technical Report
MSR-TR-95-53

Microsoft Research
Advanced Technology Division
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052

Summary

We provide a new characterization of the Bivariate normal-Wishart distribution. Let $\vec{x} = \{x_1, x_2\}$ have a non-singular Bivariate normal pdf $f(\vec{x}) = N(\vec{\mu}, W)$ with unknown mean vector $\vec{\mu}$ and unknown precision matrix W. Let $f(\vec{x}) = f(x_1)f(x_2|x_1)$ where $f(x_1) = N(m_1, 1/v_1)$ and $f(x_2|x_1) = N(m_{2|1}+b_{12}x_1, 1/v_{2|1})$. Similarly, define $\{v_2, v_{1|2}, b_{21}, m_2, m_{1|2}\}$ using the factorization $f(\vec{x}) = f(x_2)f(x_1|x_2)$. Assume $\vec{\mu}$ and W have a strictly positive joint pdf $f_{\vec{\mu},W}(\vec{\mu}, W)$. Then $f_{\vec{\mu},W}$ is a normal-Wishart pdf if and only if global independence holds, namely, $\{v_1, m_1\}\perp\{v_{2|1}, b_{12}, m_{2|1}\}$ and $\{v_2, m_2\}\perp\{v_{1|2}, b_{21}, m_{1|2}\}$ and local independence holds, namely, $\perp\{v_1^*, m_1^*\}$, $\perp\{v_{2|1}^*, b_{12}^*, m_{2|1}^*\}$ and $\perp\{v_2^*, m_2^*\}$, $\perp\{v_{1|2}^*, b_{21}^*, m_{1|2}^*\}$ (where x^* denotes the standardized r.v x and $\perp$ stands for independence). We also characterize the Bivariate pdfs that satisfy global independence alone. Such pdfs are termed Hyper-Markov laws and they are used for a decomposable prior-to-posterior analysis of Bayesian networks.

Abbreviated title. The normal-Wishart Distribution

Key words and phrases. Bayesian network, Characterization, Functional equation, Graphical model, Hyper-Markov law, Wishart distribution.

AMS Subject Classifications. Primary 62E10, 60E05; Secondary 62A15, 62C10, 39B99.

1 Introduction

Suppose $\vec{x} = \{x_1, x_2\}$ has a non-singular Bivariate normal distribution $N(\vec{\mu}, W)$ given by

$$f(\vec{x}) = (2\pi)^{-n/2}|W|^{1/2}e^{-1/2(\vec{x}-\vec{\mu})'W(\vec{x}-\vec{\mu})} \tag{1}$$

where $\vec{\mu} = (\mu_1, \mu_2)$ is a two dimensional mean vector and $W = (w_{ij})$ is an 2×2 precision matrix, namely, the inverse of a covariance matrix. We wish to determine $\vec{\mu}$ and W from a sample of values of $\vec{x}$.

The standard Bayesian approach to this statistical inference problem is to consider $\vec{\mu}$ and W as parameters, assign them a prior joint pdf and compute the posterior joint pdf of these parameters given the observed set of values. The usual prior to select is the normal-Wishart distribution, that is, the joint prior of the precision matrix W is a Wishart distribution $W(\alpha, T)$ given by

$$f_W(W) = C(\alpha, T)|W|^{(\alpha-3)/2}e^{-1/2tr\{TW\}} \tag{2}$$

where $C(\alpha, T)$ is a normalization constant, T is a symmetric matrix, and $\alpha > 1$ can be interpreted as the effective sample size (e.g., DeGroot, p. 55). The distribution $f_{\vec{\mu}|W}(\vec{\mu}|W)$ is the normal distribution $N(\vec{\mu}_0, \beta W)$ where $\beta > 0$ is another effective sample size. The joint pdf,

$$f_{\vec{\mu},W}(\vec{\mu}, W) = f_W(W)f_{\vec{\mu}|W}(\vec{\mu}|W) \tag{3}$$

is called the normal-Wishart distribution and it is selected as a prior mostly due to the fact that it is a conjugate distribution for normal distributions with unknown mean and unknown precision matrix.

A closely related variant of this approach is based on the fact that a Bivariate non-singular normal distribution $f(\vec{x})$ can be written as $f(\vec{x}) = f_1(x_1)f_{2|1}(x_2|x_1)$ where each factor is a normal distribution. More specifically, $f(x_1) = N(m_1, 1/v_1)$ and

$$f_{2|1}(x_2|x_1) = N(m_{2|1} + b_{12}x_1, 1/v_{2|1}) \tag{4}$$

where $m_1 = \mu_1$, $m_{2|1} = \mu_2 - b_{12}\mu_1$, and where μ_i is the unconditional mean of x_i, v_1 is the variance of x_1, $v_{2|1}$ is the conditional variance of x_2 given a value for x_1 and b_{12} is the regression coefficient of x_2 on x_1. The parameters $\{m_1, v_1\}$ are said to be associated with f_1 and the parameters $\{m_{2|1}, v_{2|1}, b_{12}\}$ are associated with $f_{2|1}$. Of course, the roles of x_1 and x_2 can be reversed in which case, $f(\vec{x}) = f_2(x_2)f_{1|2}(x_1|x_2)$ and the parameters, $\{m_2, v_2\}$ are associated with f_2 and $\{m_{1|2}, v_{1|2}, b_{21}\}$ are associated with $f_{1|2}$.

The first question we pose is as follows. Suppose $\{m_1, v_1\}$ are independent of $\{m_{2|1}, v_{2|1}, b_{12}\}$ and that $\{m_2, v_2\}$ are independent of $\{m_{1|2}, v_{1|2}, b_{21}\}$. These statements of independence are termed *global (parameter) independence* by [DL93] and a pdf that satisfies global independence is said to be a *Hyper-Markov law.* The term law is used to emphasize that the pdfs under study are quantifying beliefs regarding parameters; that is, a law is a "subjective" pdf regarding the parameters of an "objective" distribution. What can be said about a joint pdf of $\vec{\mu}, W$ that generates such independence statements. We show, under the assumption of a positive pdf for $\vec{\mu}, W$, that every pdf that satisfies global independence must have the form

$$f(\vec{\mu}, W) = W(\alpha, T) \cdot H(w_{12})$$

where $W(\alpha, T)$ is a Wishart pdf and H is an arbitrary function such that f is a positive pdf. This result characterizes the class of positive Bivariate Hyper-Markov laws. The importance of Hyper-Markov laws for a manageable prior-to-posterior analysis of decomposable graphical models is established in [DL93]. Geiger and Heckerman extend the prior-to-posterior analysis to any directed acyclic graphical model (called, a Bayesian network) [GH94, HG95]. In Section 4 we discuss a conjecture regarding the characterization of the n-variate Hyper-Markov laws along with its implication to the analysis of graphical models.

The second question we address is to find additional natural independence assertions that would leave the normal-Wishart distribution as the only possible prior for $\{\vec{\mu}, W\}$. Clearly, m_1 and v_1 are not independent, when a normal-Wishart prior is selected for $\{\vec{\mu}, W\}$, because the variance of m_1 depends on v_1. However, the standardized mean of x_1, denoted by m_1^*, is independent of v_1. We say that *local (parameter) independence* holds for the order x_1, x_2 if the standardized parameters associate with f_1 are mutually independent, namely m_1^* is independent of v_1^* and the standardized parameters associated with $f_{2|1}$ are mutually independent, i.e., $\{m_2^*, v_{2|1}^*, b_{12}^*\}$ are mutually independent. *Local independence* holds if it holds for the order x_1, x_2 and for the order x_2, x_1. The main result of this article is that if local and global independence hold, and assuming a positive pdf $f(\vec{\mu}, W)$, then $f(\vec{\mu}, W)$ must be a normal-Wishart distribution. Therefore, local and global independence characterize the positive Bivariate normal-Wishart distribution. We believe that the positivity assumption is actually redundant.

The mathematical tool we use for the proof of the normal-Wishart characterization is the theory of functional equations [Ac66]. In section 2 we develop the needed functional equation and use results of [Ja86] to prove that every positive measurable solution of it has infinitely many derivatives. In sections 3 through 5 we gradually solve this equation and in Section 6 we outline a related work [GH95] and possible extensions along with their statistical application.

2 The functional equation

Let $\vec{x} = \{x_1, x_2\}$ have a non-singular Bivariate normal pdf $f(\vec{x}) = N(\vec{\mu}, W)$. If we write $f(\vec{x}) = f(x_1)f(x_2|x_1)$ where $f(x_1) = N(m_1, 1/v_1)$ and $f(x_2|x_1) = N(m_{2|1} + b_{12}x_1, 1/v_{2|1})$, then the following well known relationships are satisfied:

$$w_{11} = \frac{1}{v_1} + \frac{b_{12}^2}{v_{2|1}} \qquad w_{12} = -\frac{b_{12}}{v_{2|1}} \qquad w_{22} = \frac{1}{v_{2|1}} \qquad m_1 = \mu_1 \qquad m_{2|1} = \mu_2 - b_{12}\mu_1 \tag{5}$$

Similar equalities hold when $f(\vec{x})$ is written as $f_2(x_2)f_{1|2}(x_1|x_2)$. Note that the transformation between $\{\vec{\mu}, W\}$ and $\{m_1, v_1, m_{2|1}, v_{2|1}, b_{12}\}$ is one to one and onto as long as W is the inverse of a covariance matrix and the variances $v_1, v_{2|1}$ are positive. The Jacobian of this transformation is given by,

$$\left|\frac{\partial w_{11}, w_{12}, w_{22}, \mu_1, \mu_2}{\partial v_1, v_{1|2}, b_{12}, m_1, m_{2|1}}\right| = v_1^{-2} v_{2|1}^{-3} \tag{6}$$

The statement $x \perp y$ denotes that x and y are independent and $\perp\{x, y, z\}$ denotes that x, y and z are mutually independent. For a random variable x, let x^* denote its corresponding standardized random variable, i.e., the centralized x divided by its standard deviation.

We can now state the characterization theorem.

Theorem 1 *Let $\vec{x} = \{x_1, x_2\}$ have a non-singular Bivariate normal pdf $f(\vec{x}) = N(\vec{\mu}, W)$ where $\vec{\mu}$ and W are unknown. Assume $\vec{\mu}$ and W have a strictly positive joint pdf $f_{\vec{\mu},W}(\vec{\mu}, W)$. Then $f_{\vec{\mu},W}$ is a normal-Wishart pdf if and only if global independence holds, namely, $\{v_1, m_1\} \perp \{v_{2|1}, b_{12}, m_{2|1}\}$ and $\{v_2, m_2\} \perp \{v_{1|2}, b_{21}, m_{1|2}\}$ and local independence holds, namely, $\perp\{v_1^*, m_1^*\}$, $\perp\{v_{2|1}^*, b_{12}^*, m_{2|1}^*\}$ and $\perp\{v_2^*, m_2^*\}$, $\perp\{v_{1|2}^*, b_{21}^*, m_{1|2}^*\}$.*

The pdf $f_{\vec{\mu},W}$ can be written, by a change of variables, in terms of $v_1, v_{2|1}, b_{12}, m_1, m_{2|1}$ as well as in terms of $v_2, v_{1|2}, b_{21}, m_2, m_{1|2}$, using the Jacobian given by Equation 6. Consequently, it is immediate to verify that global and local independence are satisfied. For the converse, since both representations must be equal, and using the global independence assumption made by Theorem 1, we get the following equality each side of which is equal to $f_{\vec{\mu},W}$,

$$\frac{1}{v_1^2 v_{2|1}^3} f_1(m_1, v_1)\ f_{2|1}(m_{2|1}, b_{12}, v_{2|1}) = \frac{1}{v_2^2 v_{1|2}^3}\ f_2(m_2, v_2)\ f_{1|2}(m_{1|2}, b_{21}, v_{1|2}) \tag{7}$$

where f_1, $f_{2|1}$, f_2, and $f_{1|2}$ are arbitrary pdfs, and, from Equation 5,

$$v_2 = v_{2|1} + v_1 b_{12}^2 \qquad b_{21} = \frac{b_{12}v_1}{v_2} \qquad v_{1|2} = \frac{v_{2|1}v_1}{v_2} \qquad m_2 = m_{2|1} + b_{12}m_1 \qquad m_{1|2} = m_1 - b_{21}m_2 \tag{8}$$

This equality is in fact a functional equation that encodes the assumption of global independence. By solving it, we obtain all pdfs that satisfy global independence. At the second step we incorporate the assumption of local independence into the solutions obtained to show that $f_{\vec{\mu},W}$ must be a normal-Wishart pdf. Note that we may regard $v_1 > 0, v_{2|1} > 0, b_{12} \neq 0, m_1, m_{2|1}$ as free variables and all other variables are defined by Equation 8.

Methods for solving functional equations such as Eq. 7, that is, finding all functions that satisfy them under different regularity assumptions, are discussed in [Ac66]. We use the following technique. We take repeated derivatives of a logarithmic transformation of Eq. 7 and obtain a differential equation the solution of which after appropriate specialization is the general solution of Eq. 7 (Aczél, 66, Section 4.2, "Reduction to differential equations").

The reduction to differential equations is justified by the results in [Ja86]. Járai has extensively investigated the following type of functional equations:

$$f(t) = \sum_{i=1}^{n} h_i(t, y, f_i(g_i(t, y))) \tag{9}$$

where $f, f_1, \ldots, f_n$ are unknown functions, $h_1, \ldots, h_n, g_1, \ldots, g_n$ are known functions satisfying some regularity conditions and all variable and function values may be vectors. By taking the logarithm of Equation 7, which is justified since we assume all pdfs are positive, we obtain a functional equation of the form given by Equation 9. Járai showed, among other results, that every measurable solution of Equation 9 must have infinitely many derivatives (Theorems 3.3, 5.2, 7.2 in [Ja86]). Since Equation 7 satisfies the needed regularity conditions, we may conclude that any positive pdf that solves it must have indefinitely many derivatives (since a pdf is Lebesgue integrable and thus measurable). We conjecture, though, that every measurable solution of Equation 7 which is positive on a measurable subset of a positive measure (i.e., any pdf), must be positive and so we believe the assumption of positiveness is actually redundant. The importance of Járai's results is that in the process of solving Equation 7 one may take as many derivatives as is found needed. Further examples of the applicability of Járai's results to characterization problems in statistics are discussed in [GH95] along with an example where they fail.

3 The Bivariate equation with a fixed precision matrix

We shall first assume that the variances v_1 and $v_{2|1}$ in Equation 7 are fixed and that b_{12} is fixed and is not zero. Thus, by renaming variables and functions, Equation 7 can be stated as follows:

$$\tilde{f}(m)\tilde{g}(n) = \tilde{h}(n + wm)\tilde{k}(m - \frac{yw}{x}(n + wm)) \tag{10}$$

where $x = z + yw^2, w \neq 0, y > 0$, and $z > 0$. The free variables are m and n which correspond to m_1 and $m_{2|1}$ while x, y, z and w are fixed constants which correspond to v_2, v_1, $v_{2|1}$ and b_{12}, respectively. Let $u = m - \frac{yw}{x}(n + wm)$. Let $a(t) = \log \tilde{a}(t)$ for every function $\tilde{a}$.

Taking the logarithm and then a derivative once wrt (with respect to) m and once wrt n yields the following two equations, respectively:

$$f'(m) = wh'(n + wm) + \frac{z}{x}k'(u) \; ; \qquad g'(n) = h'(n + wm) - \frac{yw}{x}k'(u)$$

By taking a derivative wrt n and m of these equations, we get, $f''(m) = w^2 h''(n + wm) + \frac{z^2}{x^2}k''(u)$ and

$$g''(n) = h''(n + wm) + \frac{y^2 w^2}{x^2}k''(u) \; ; \qquad 0 = wh''(n + wm) - \frac{zyw}{x^2}k''(u)$$

Consequently, using $x = z + yw^2$, we obtain,

$$f''(m) = \frac{z}{y}g''(n) \tag{11}$$

for every m and n. Since z and y are held fixed, $f''(m)$ and $g''(n)$ must both be constant functions and therefore by integrating twice and then taking the exponent we get,

$$\tilde{f}(m) = e^{c_1 m^2 + c_2 m + c_3}$$

which implies that $\tilde{f}$ is a normal pdf. Similarly $\tilde{g}$ is normal and by symmetric arguments $\tilde{h}$ and $\tilde{k}$ are normal pdfs as well. Thus, we may conclude that for a fixed objective precision matrix, global parameter independence implies that $\{m_1, m_{2|1}\}$ have independent normal pdfs and so do $\{m_2, m_{1|2}\}$ have. It follows that $\{m_1, m_2\}$ have a joint normal pdf.

One may regard this result as a characterization of a normal pdf, however, this result is by no means new since a more general characterization of a normal pdf is well known as the Skitovich-Darmois theorem; Suppose $X_1, \ldots X_n$ are independent random variables and let $L_1 = \sum \alpha_i X_i$ and $L_2 = \sum \beta_i X_i$ be two linear statistics where α_i and β_i are constant coefficients. If L_1 and L_2 are independent, then the random variables X_j for which $\alpha_j \beta_j \neq 0$ are all normal. (This result is described in [KLR73, Theorem 3.1.1]) along with its history). This theorem is applicable by choosing $X_1 = m_1$, $X_2 = m_{2|1}$, $L_1 = m_2$ and $L_2 = m_{1|2}$ (See Eq. 8).

The new twist emphasized herein is that global parameter independence implies a joint subjective normal pdf for the unknown means of an objective normal pdf with a fixed precision matrix. Furthermore, we demonstrate the powerful and elementary tools of functional equations. Járai's results can now be called upon to prove that any positive measurable function that satisfies Equation 7 must have infinitely many derivatives and therefore the steps we have taken to solve this equation are justified.

4 The fixed means Bivariate equation

We shall now assume that the means m_1 and $m_{2|1}$ in Equation 7 are fixed. Consequently, we obtain the following weaker version of Theorem 1 which characterizes the Wishart distribution.

Theorem 2 *Let $\vec{x} = \{x_1, x_2\}$ have a non-singular Bivariate normal pdf $f(\vec{x}) = N(\vec{\mu}, W)$ where $\vec{\mu}$ is known and W is unknown. Assume W has a strictly positive joint pdf $f_W(W)$. Then f_W is a Wishart pdf if and only if global independence holds, namely, $\{v_1\} \perp \{v_{2|1}, b_{12}\}$ and $\{v_2\} \perp \{v_{1|2}, b_{21}\}$ and local independence holds, namely, $\perp\{v^*_{2|1}, b^*_{12}\}$ and $\perp\{v^*_{1|2}, b^*_{21}\}$.*

For known $\vec{\mu}$ Equation 7 can be stated as follows:

$$\tilde{f}(y)\tilde{g}(z,w) = \tilde{h}(x)\tilde{k}(\frac{yz}{x}, \frac{yw}{x}) \tag{12}$$

where $x = z + yw^2$, $w \neq 0, y > 0$, and $z > 0$. Recall that $a(y) = \log \tilde{a}(y)$ for every function $\tilde{a}$. Also let $g_i(z,w)$ denote the first derivative wrt argument i of $\log \tilde{g}(z,w)$ and let $g_{ij}(z,w)$ denote the mixed derivative.

Taking the logarithm Eq. 12 and then a derivative once wrt z, once wrt w, and once wrt y yields the following three equations, respectively:

$$g_1(z,w) - h'(x) = \frac{y^2w^2}{x^2}k_1(\frac{yz}{x}, \frac{yw}{x}) - \frac{yw}{x^2}k_2(\frac{yz}{x}, \frac{yw}{x}) \tag{13}$$

$$g_2(z,w) - 2ywh'(x) = -\frac{2y^2zw}{x^2}k_1(\frac{yz}{x}, \frac{yw}{x}) + \frac{z - yw^2}{x^2}yk_2(\frac{yz}{x}, \frac{yw}{x}) \tag{14}$$

$$f'(y) - w^2h'(x) = \frac{z^2}{x^2}k_1(\frac{yz}{x}, \frac{yw}{x}) + \frac{zw}{x^2}k_2(\frac{yz}{x}, \frac{yw}{x}) \tag{15}$$

Solving for $k_1(\frac{yz}{x}, \frac{yw}{x})$ and $k_2(\frac{yz}{x}, \frac{yw}{x})$ using Eqs. 13 and 14, plugging the result into Equation 15 and simplifying using $x = z + yw^2$, yields,

$$z^2g_1(z,w) + wzg_2(z,w) = y^2w^2f'(y) - x^2h'(x) \tag{16}$$

Taking a derivative wrt z yields

$$\frac{d}{dz}\left[z^2g_1(z,w) + wzg_2(z,w)\right] = 2xh'(x) + x^2h''(x) \tag{17}$$

Consequently, since the right hand side of Equation 17 is a function of $y + zw^2$ while the left hand side is not a function of y, it follows that both the right and left hand sides of this equation are equal to a constant, say c_1. We are left with the following differential equation:

$$x^2h''(x) + 2xh'(x) = c_1 \ ; \qquad \left[x^2h'(x)\right]' = c_1 \tag{18}$$

Consequently, and due to the symmetric roles of f and h, we have,

$$h'(x) = \frac{c_1}{x} + \frac{c_2}{x^2} \; ; \qquad f'(y) = \frac{c_3}{y} + \frac{c_4}{y^2} \tag{19}$$

Now from Equation 19 we get, after integration and exponentiation,

$$\tilde{h}(x) \propto x^{-c_1} e^{-\frac{c_2}{x}} \; ; \qquad \tilde{f}(y) \propto y^{-c_1} e^{-\frac{c_4}{y}} \tag{20}$$

Recall that x and y correspond to the unconditional variances v_2 and v_1. Consequently, we have shown that global parameter independence alone implies that v_1 and v_2 have an inverse Gamma distribution given by Equation 20. One may in fact consider this result as a characterization of the Gamma distribution. Of course, given a Wishart distribution, it is well known that the marginal distribution for v_1 and v_2 are inverse Gamma, but there are possibly many other pdfs with such marginal distributions. We proceed to show that there are further restrictions implied by Equation 12. To do so we find the general solution for $g(z, w)$.

The solution for $g(z, w)$ is determined by the partial differential equation obtained by plugging Equation 19 into Equation 16, which yields,

$$z^2 g_1(z, w) + wz g_2(z, w) = -c_1 x - c_2 + c_3 y w^2 + c_4 w^2$$

and, since y must cancel on the right hand side, $c_1 = c_3$, it finally reduces to,

$$z^2 g_1(z, w) + wz g_2(z, w) = -c_1 z - c_2 + c_4 w^2 \tag{21}$$

The corresponding homogeneous differential equation, after a change of variables $s = z$ and $t = w/z$, becomes, $s^2 g_s(s, t) = 0$ $(s > 0)$, which yields by integration wrt s that $g(s, t) = \tilde{H}(t)$ where $\tilde{H}$ is an arbitrary differentiable function. One solution to this differential equation, which can be verified by substituting it into this equation, is given by,

$$g_p(z, w) = c_1 \log \frac{1}{z} + \frac{c_2}{z} + \frac{c_4 w^2}{z} + c_8$$

Consequently, the general solution after exponentiating is given by

$$\tilde{g}(z, w) \propto z^{-c_1} e^{-\frac{c_2}{z}} e^{\frac{c_4 w^2}{z}} \tilde{H}(w/z) \tag{22}$$

We can further rewrite Equation 22 as follows,

$$\tilde{g}(z, w) \propto z^{-c_1} e^{\frac{-c_5}{z}} e^{-\frac{(w-w_0)^2}{2c_6 z}} \hat{H}(w/z) \tag{23}$$

where w_0 is the expectation of w. Thus the general solution for $f_W(W)$ is given by,

$$f_W(W) \propto y^{-c_1} e^{-\frac{c_4}{y}} z^{-c_1} e^{\frac{-c_5}{z}} e^{-\frac{(w-w_0)^2}{2c_6 z}} \hat{H}(w/z) \tag{24}$$

Rewriting y, z and w in terms of W and using the fact that v_2 is independent of $\{v_{1|2}, b_{21}\}$ yield $f_W(W) = W(\alpha, T) \cdot H(w_{12})$. where $H(w_{12}) = \hat{H}(-w_{12})$. This is the most general type of pdf that satisfies global independence. When H is constant, the resulting pdf is a Wishart. We are not aware of other known Bivariate pdfs that can be written in this form. We shall now show that if local independence holds H must be constant and the Wishart pdf characterization is completed.

Let $v = \sqrt{z}$ and $t = (w - w_0)/v$. That is, t is the standardized regression coefficient. Using this notation, Equation 23 has the following functional form,

$$\tilde{g}(v,t) = a(v)b(t^2)H(\frac{w_0 + vt}{v^2}) \tag{25}$$

Local independence implies $\tilde{g}(v,t) = d(v)e(t)$, and therefore, Equation 25 reduces to

$$H(\frac{w_0 + vt}{v^2}) = \tilde{G}(v)\tilde{K}(t) \tag{26}$$

Taking the logarithm yields,

$$L(\frac{w_0 + vt}{v^2}) = G(v) + K(t) \tag{27}$$

Taking a derivative once wrt t and once wrt v yields

$$\frac{1}{v}L'(\frac{w_0 + vt}{v^2}) = K'(t) \ ; \qquad -(\frac{2w_0}{v^3} + \frac{t}{v^2})L'(\frac{w_0 + vt}{v^2}) = G'(v), \tag{28}$$

and therefore,

$$-(\frac{2w_0}{v^2} + \frac{t}{v})K'(t) = G'(v) \tag{29}$$

By setting two distinct values t_1 and t_2, we get,

$$(\frac{2w_0}{v^2} + \frac{t_1}{v})K'(t_1) = (\frac{2w_0}{v^2} + \frac{t_2}{v})K'(t_2) \tag{30}$$

Thus either $K'(t_1) = K'(t_2) = 0$ or $w_0 = 0$ in which case $t_1K'(t_1) = t_2K'(t_2)$. In the first case, due to Equation 28, $L'(t) = 0$. Consequently, $L(t)$ is constant and so is $H(t)$. In the second case, $K'(t) = c/t$ for some constant c, thus, from Equation 28 (with $w_0 = 0$), $L'(x) = c/x$ and so $H(x) = c_2x^c$.

Recalling the probabilistic interpretation of Equation 24 where $z = v_{2|1}$ and $w = b_{12}$ we observe that the conditional pdf of w given $v_{2|1}$ is a generalized normal pdf given by,

$$f(w|v_{2|1}) \propto w^c e^{-\frac{(w-w_0)^2}{2c_6 v_{2|1}}}$$

Thus, for w_0 to be the expectation of w, c must equal zero, implying that $H(x)$ is constant.

5 The full Bivariate equation

In this section we prove Theorem 1 in its full extent. By renaming variables and functions, Equation 7 can be stated as follows:

$$\tilde{f}(m,y)\tilde{g}(n,z,w) = \tilde{h}(n+wm,x)\tilde{k}(u,\frac{yz}{x},\frac{yw}{x}) \tag{31}$$

where $x = z + yw^2, y > 0$, and $z > 0$ and where $u = \frac{z}{x}m - \frac{yw}{x}n$. Using the same technique as in Section 4 we obtain the following equality between the second partial derivatives of f and g wrt their first argument,

$$yf_{11}(m,y) = zg_{11}(n,z,w) \tag{32}$$

(As in Equation 11). Thus each side must equal a constant and by integration, we get

$$yf_1(m,y) = 2cm + a(y)\ ; \qquad zg_1(n,z,w) = 2cn + b(z,w) \tag{33}$$

Similarly,

$$xh_1(n+wm,x) = 2e\cdot(n+wm) + r(x)\ ; \qquad \frac{yz}{x}k_1(u,\frac{yz}{x},\frac{yw}{x}) = 2eu + q(\frac{yz}{x},\frac{yw}{x})$$

Taking the logarithm and then a derivative wrt m of Equation 31 yields

$$f_1(m,y) = wh_1(n+wm,x) + \frac{z}{x}k_1(u,\frac{yz}{x},\frac{yw}{x}) \tag{34}$$

Plugging the expressions for f_1, h_1 and k_1 and grouping all terms that contain m yield that $e = c$, and, using $x = z + yw^2$, we get,

$$a(y) = ywR(x) + q(\frac{yz}{x},\frac{yw}{x}) \tag{35}$$

where $R(x) = r(x)/x$. Taking a derivative once wrt y, once wrt z once wrt w yields the following three equations:

$$a'(y) = wR(x) + yw^3R'(x) + \frac{z^2}{x^2}q_1(\frac{yz}{x},\frac{yw}{x}) + \frac{wz}{x^2}q_2(\frac{yz}{x},\frac{yw}{x}) \tag{36}$$

$$0 = yR'(x)w + \frac{y^2w^2}{x^2}q_1(\frac{yz}{x},\frac{yw}{x}) - \frac{yw}{x^2}q_2(\frac{yz}{x},\frac{yw}{x})$$

$$0 = yR(x) + 2y^2w^2R'(x) - \frac{2y^2wz}{x^2}q_1(\frac{yz}{x},\frac{yw}{x}) + y\frac{z-yw^2}{x^2}q_2(\frac{yz}{x},\frac{yw}{x})$$

The last two equations yield,

$$wq_2(\frac{yz}{x},\frac{yw}{x}) = 2wx^2R'(x) + xwR(x) \tag{37}$$

$$yw^2 q_1(\frac{yz}{x}, \frac{yw}{x}) = wx^2 R'(x) + wxR(x) \tag{38}$$

Plugging these equations back into Equation 36 yields, using $x = z + yw^2$, $w \neq 0$,

$$ya'(y) = \frac{x}{w} R(x) + \frac{x^2}{w} R'(x) \tag{39}$$

Setting $w = 1$ yields that both sides of Equation 39 must equal a constant, say C. Thus $ya'(y) = C$ and

$$Cw = xR(x) + x^2 R'(x) \tag{40}$$

Now setting $y = 1$ in Equation 40 yields that $C = 0$. Thus $a(y) = a$ where a is a constant. Consequently, from Equation 40 we get $R'(x)/R(x) = -1/x$ and therefore, $R(x) = r/x$ where r is a constant. Thus, from Equation 35, $q(s,t) = a - rt$.

Plugging these solutions into Equation 31 yields an equation where m and n cancel out, and y, z and w as free variables. This is precisely the functional equation given by Equation 12. Thus, the joint distribution for the precision matrix is given by Equation 24 and it is a Wishart pdf if local independence is assumed. By integrating Equation 33, using $a(y) = a$ we immediately obtain that the conditional distribution of m is normal, and similarly for n.

Concluding Remarks

The notion of local and global parameter independence has also been studied in the context of Bayesian estimation of a joint pdf $p(s,t)$ from a sample of pairs of values when s and t are two discrete random variables. An analogous result to Theorem 1 has been established in [GH95] which characterizes the Dirichlet distribution. Let θ_{ij}, $1 \leq i \leq k, 1 \leq j \leq n$, be positive random variables that sum to unity. Define $\theta_{i\cdot} = \sum_{j=1}^{n} \theta_{ij}$, $\theta_{I\cdot} = \{\theta_{i\cdot}\}_{i=1}^{k-1}$, $\theta_{j|i} = \theta_{ij}/\sum_j \theta_{ij}$, and $\theta_{J|i} = \{\theta_{j|i}\}_{j=1}^{n-1}$. Geiger and Heckerman (1995) show that if and only if $\{\theta_{I\cdot}, \theta_{J|1}, \ldots, \theta_{J|k}\}$ are mutually independent and $\{\theta_{\cdot J}, \theta_{I|1}, \ldots, \theta_{I|n}\}$ are mutually independent (where $\theta_{\cdot J}$ and $\theta_{I|j}$ are defined analogously), and assuming strictly positive pdfs, the pdf of θ_{ij} is Dirichlet.

It is interesting to observe that local and global parameter independence characterize many known conjugate sampling distributions including, the Dirichlet, the Bivariate normal (Section 3), the Gamma distribution (Section 4) and the Bivariate Wishart distribution. We conjecture that global and local parameter independence also characterize the n-dimensional Wishart distribution. In the case of Dirichlet distributions, the n-variate characterization follows from the Bivariate characterization without solving additional functional equations [GH95, HGC94]. However, for the normal-Wishart characterization the situation looks more complex since new functional equations are introduced as the dimensionality increases.

In the study of decomposable graphical models and other Bayesian networks the assumption of local independence can be bypassed without serious complications, however the assumption of global independence is indispensable for achieving a manageable prior-to-posterior analysis [DL93, GH94]. Thus it is important to characterize the set distributions that satisfy global independence alone and to provide specific examples of such distributions that can be used as prior distributions for model selection of graphical models. This article provides a step towards this goal.

References

[Ac66] J. Aczél, *Lectures on functional equations and their applications*, Academic Press, New York, 1966.

[DL93] P. Dawid and S. Lauritzen, Hyper Markov laws in statistical analysis of decomposable graphical models, *Annals of Statistics*, 21:1272–1317, 1993.

[De70] M. DeGroot, *Optimal statistical decisions*, McGraw-Hill, New York.

[GH95] D. Geiger and D. Heckerman, A characterization of the Dirichlet distribution through local and global parameter independence. Submitted to Annals of Statistics, February 1995.

[GH94] D. Geiger and D. Heckerman, Learning Gaussian Networks, Proceedings of Tenth Conference on Uncertainty in Artificial Intelligence, Seattle, WA, Morgan Kaufmann, July 1994, 235–243.

[HGC94] D. Heckerman, D. Geiger and D. Chickering, Learning Bayesian Networks: The combination of knowledge and statistical data, Proceedings of Tenth Conference on Uncertainty in Artificial Intelligence, Seattle, WA, Morgan Kaufmann, July 1994, 293–301.

[HG95] D. Heckerman and D. Geiger, Learning Bayesian Networks: A unification for discrete and Gaussian domains. Proceedings of Eleventh Conference on Uncertainty in Artificial Intelligence, Montreal, Canada, Morgan Kaufmann, August 1995.

[Ja86] A. Járai, On regular solutions of functional equations, *Aequationes Mathematicae*, 30, 21–54, 1986.

[KLR73] Kagan A.M., Rao C.R. and Y.V. Linnik, *Characterization problems in mathmatical statistics*, John Wiley and Sons, 1973.

Permanent address:

Dan Geiger, Computer Science Department, Technion, Haifa 32000, Israel.
David Heckerman, Microsoft Research, Redmond WA, 98052, USA.